Functional Safety Assessment in Autonomous Vehicles

Akshay Kumar Shastry

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Engineering

Haibo Zeng, Chair
Michael S. Hsiao
Cameron D Patterson

May 09, 2018
Blacksburg, Virginia

Keywords: Functional Safety, Autonomous Vehicles, Electronic Control Software Architecture, Influence Diagrams

Functional Safety Assessment in Autonomous Vehicles

Akshay Kumar Shastry

(ABSTRACT)

Autonomous vehicles (AVs) are a class of safety-critical systems that are capable of decision-making and operate with little or no human intervention. For such complex systems designed to function in diverse operational domains such as rain, snow, freeway, urban roads, etc., system safety is paramount. Management of the system's safety throughout its life-cycle, from the conceptualization stage to the end of the lifecycle, is of primary importance. We describe a revision of functional safety standard ISO 26262 to support autonomous vehicles and the underlying electronic/electrical control architecture. There is a need to modify the Automotive Safety Integrity Levels (ASILs) defined in the ISO 26262 as "Controllability", a factor in determining an ASIL, is no longer applicable; the driver is no longer in a position to control the vehicle. The vehicle has taken over the responsibility of evaluating the environment and determines its next course of action to complete its current mission. These decisions have a tremendous impact on the overall safety of the system during a hazardous event and can be the difference between a successful journey and a traffic incident. To better enable the designers of such systems, we introduce a new method to assess functional safety and derive safety goals, which are the top level safety requirement. We present a new metric-Risk Mitigation Factor to assess the decision making capability of the vehicle and to replace controllability in the ASIL definition. The case study presented highlights the advantages of using the introduced metric in defining safety goals for the autonomous vehicle.

Functional Safety Assessment in Autonomous Vehicles

Akshay Kumar Shastry

(GENERAL AUDIENCE ABSTRACT)

Autonomous vehicles (AVs) are changing the way we perceive mobility and transportation. AVs are soon to be a part of everyday life, from giving you a ride to the office to taking children to the dentist. All the possible benefits of AVs are attainable if the systems designed are safe for use. Safety in AVs is the primary challenge in design and development. It is crucial to incorporate the principles of safety in system design from the beginning of the inception phase to the end of the lifecycle of the vehicle. The challenges for ensuring safety in AVs are enormous, from implementing the correct operation for a system to assuring that system behavior is safe in the presence of a malfunction; the scale and complexity of the systems drive the safety requirements. In the work presented, we focus on the functional safety of the underlying electrical/ electronic architecture of the vehicle, describing a revision of the automotive functional safety standard ISO 26262 for AV development. We propose to leverage the decision-making capabilities of the vehicle to assure safety in a hazardous situation.

Dedication

I dedicate this to my parents.

Acknowledgments

I would like to thank my advisor Dr. Haibo Zeng, for his constant support, patience, guidance and for providing the wonderful opportunity to work on challenging areas of research throughout my masters. I would like to thank my family for their understanding, support and motivation. Finally, I would like to thank my friends and labmates who have been a part of my journey at Virginia Tech.

Contents

List of Figures

List of Tables

AV - Autonomous Vehicle

ASIL - Automotive Safety Integrity Level

E/E - Electrical and Electronic

FO - Fail Operational

FS - Fail Safe

F-Sil - Fail Silent

FMEA - Failure Mode Element Analysis

FTA - Fault Tree Analysis

GSN - Goal State Notation

ID - Influence Diagram

HARA - Hazard and Risk Analysis

IEEE - Institute of Electrical and Electronic Engineers

LKAS - Lane Keeping Assistance

NHTSA - National Highway Traffic Safety Administration

ODD - Operational Design Domain

SAE - Society of Automotive Engineers

STPA - System Theoretic Process and Analysis

Chapter 1

Introduction

In the current chapter, Section 1.1 discusses the motivation behind the work and introduces the context of the problem statement. Section 1.2 provides a background to the terminology used in the standard ISO 26262 [11]. Section 1.3 illustrates the functional architecture of an AV and Section 1.4 discusses associated concepts from reliability engineering. Section 1.5 introduces the problem statement and Section 1.6 highlights the idea and the contribution behind the current work. Section 1.7 summarizes the related work. Section 1.8 presents the structure of the report.

1.1 Motivation

Autonomous vehicles (AV) are expected to become a part of everyday life shortly with original equipment manufacturers such as General Motors, Ford, and Tesla testing the technology extensively [32] alongwith non-traditional companies such as Waymo and Uber. Autonomous vehicles are expected to reduce the number of casualties due to hazardous traffic incidents [19]. They are projected to change the way society perceives mobility. To completely utilize the benefits and the advantages of autonomous vehicles, we need to ensure the safety of the system and those who use it. The key benefit of an autonomous vehicle also proves to be a key concern: There is no driver, there are only users. Since the driver can no longer be relied upon to take evasive maneuvers in a hazardous situation, the system safety needs to

be assured in design during the development phase and its capability to handle an emergency situation requires assessment. A potential fault or error in the system can lead to dangerous driving situations in which passengers or pedestrians are harmed if appropriate redundancies and contingencies are not in place. Safety is an inherent system quality. We need to incorporate safety in design from the concept phase for any AV, and the management of system safety needs to be determined until the end of the vehicle's intended service lifecycle. With AVs we need not just concern ourselves with correct operation under normal conditions, but also need to pay close attention to the vehicle's behavior when a component or subsystem malfunctions. Designing for safety is a non-trivial task as these systems are resource constrained and driven by strict time-to-market constraints. Modeling every possible driving scenario and determining the vehicle's behavior exhaustively may not be possible. The economics of the design also comes into play, overdesigning a car may make it economically unfeasible to produce or sell. There is a need to efficiently design the system for safety without any compromise while dealing with the constraints.

ISO 26262 is the standard for functional safety in road vehicles [11]. ISO 26262 defines safety lifecycle of the vehicle and management of safety-related systems in the electrical and/or electronic (E/E) architecture of the vehicle. Functional safety is the overall safety of the system, which depends on the system/component operating correctly in response to the inputs, errors, faults and the operating conditions. A safety goal is a top-level safety requirement derived from the hazard and risk assessment (HARA) of the system. The safety goals and the allocation of the Automotive Safety Integrity Levels (ASILs) prescribed in the standard drive the process of development. The integrity requirements of hardware and software platform technologies and their development ensure that specific error handling capabilities are in-built. The safety lifecycle plays a vital role in the design and has to be incorporated from the inception phase before the development.

The international society of automotive engineers (SAE International) defines six levels of autonomy in vehicles [24] as shown in Figure 1.1. A level 5 car is completely autonomous where all the functions of a driver in all driving domains are taken over by the system. The standard is insufficient for assessing the safety requirements of a level 5 AV as it assumes that the driver is in the loop, and is capable of handling any potentially hazardous situation. There is a need to redefine the metrics that assess the safety requirements of the vehicle.

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	**No Automation**	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	**Driver Assistance**	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	**Partial Automation**	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* ("system") monitors the driving environment**						
3	**Conditional Automation**	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	**High Automation**	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	**Full Automation**	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

Figure 1.1: Summary of Levels of Autonomy defined by SAE International in J3016

The current functional safety concept is based on acceptable risk for a hazard. Active safety mechanisms need to be incorporated into the system to deal with unacceptable levels of risk. In ISO 26262, Automotive Safety Integrity Level (ASIL) is defined as a product of severity

of harm (Severity), probability of the driving situation (Exposure) and whether in a driving situation the vehicle is controllable by the driver (Controllability). Figure 1.2 summarizes the ISO 26262 process in four main steps, 1) Item Definition, 2) Hazard and Risk Analysis, 3) ASIL Assignment and, 4) Determination of safety goals.

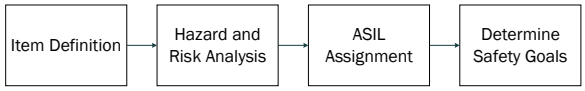

Figure 1.2: Summary of current ISO 26262 practice

The term "Controllability" is no longer applicable as a factor to determine the ASIL levels for any item, as the users of the vehicle cannot take any actions to reduce the possibility of harm in a hazardous situation. Severity and Exposure are still applicable as the extent of harm, and the probability of a particular driving situation does not change. Controllability would be suitable if the system behavior modeled after human drivers behaves exactly like a human driver would behave and reacts to a given hazardous situation appropriately. Since this is not the case, the current definition of ASIL requires modification. Allocating Controllability as C3 or uncontrollable for automated functions seems feasible, but it leads to a very conservative allocation of ASILs. A conservative allocation of ASILs may lead to over-designing the system due to the safety requirements, potentially increasing the cost of development, testing, and production phases substantially.

The safety goals and functional safety concept for vehicles are defined assuming that the driver is the redundant factor in a hazardous situation and is capable of mitigating the hazard through some action/maneuver. In a fully autonomous vehicle, the driver is a user of a service provided by the vehicle, rather than the entity responsible for it. The driver is not a part of any decision making or environmental perception, these responsibilities have been taken over by the vehicle. Therefore, defining a quantifiable metric that can be used

to evaluate this decision making capability of the vehicle in the absence of driver becomes necessary. Without such a defined metric, evaluating the functional safety and identifying the set of safety goals for an autonomous vehicle is difficult. Safety of intended function (SOTIF) is another concern in AV development. Safety of intended function is where the system fails to perform as expected in the absence of a fault due to some design oversight resulting in a hazardous situation. A few examples are – when a LiDAR falsely detects an obstacle due to accumulated dirt on the sensing lens or when a lane recognition system fails to recognize the merging of lanes due to faded lane markings. These faults are a combination of systematic faults that are inherent in the design of the algorithm and the errors in interpretation due to an unaccounted driving situation. These cases are not covered explicitly by ISO 26262. Any new approach will have to incorporate safety concepts to cover these types of operational situations as well.

1.2 ISO 26262 - A Short Background

ISO 26262 operates on the idea of unreasonable risk. The standard is divided into ten parts and relies on hazard and risk analysis to analyze potential hazards. The standard advocates for the V-model of development for both hardware and software of the system. The interested reader may refer to the standard for further details. The following terms and definitions are defined in part 1 of ISO 26262 [11]. :

- Harm : Physical injury or damage.
- Severity (S): Measure of extent of harm to an individual.
- Exposure (E) : State of being in an operational situation that can be hazardous if coincident with the failure mode in analysis.

- Controllability (C) : The ability to avoid harm or damage through timely reactions of the persons involved, possibly with the support of external measures. It usually depends on the possible actions of the driver or the pedestrians involved to avoid harm.

- Item : System or array of systems to implement a function at system level.

- Hazard : Potential source of harm caused by the malfunctioning behavior of the item.

- Risk : combination of probability of occurrence of harm (P_H) and severity of that harm (S) in a specific situation.

- Safety Goal : Top level safety requirement of the system from a hazard and risk analysis.

- Safe state : operating mode of an item without unreasonable risk.

- Fault tolerant time interval (FTTI) : Time span in which a fault can be present before the possible occurrence of a hazardous event (Figure 1.3).

- Fault reaction Time (FRT): Time span between detection of a fault to reaching a safe state (Figure 1.3).

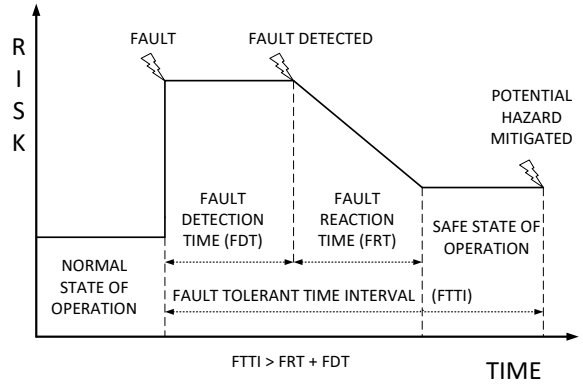

Figure 1.3: Fault Tolerant Time Interval

The part 3 of the ISO 26262 deals with the concept phase. The standard does not address the performance aspect of the electronic subsystems, it focuses on the hazards caused by the faulty behavior in the subsystems. The part 3 of ISO 26262 in several clauses outlines the definition of an item, the hazard and risk analysis (HARA), assignment of ASIL, the derivation of the safety goals, and the functional safety concept. The functional and non-functional requirements as well as the boundary of the item and its interfaces are parts of the item definition. For the defined item, we identify and categorize the possible hazards that can experience due a malfunction. These are defined as hazardous events (HEs). We then assign ASILs to the HEs. The assignment of ASILs is followed by the derivation of the appropriate safety goal which is the top-level safety requirement. ASILs are assigned using severity, exposure, and controllability shown in Tables 1.1, 1.2 and 1.3 respectively.

S0	S1	S2	S3
No injuries	Light to Moderate injuries	Severe life threatening injuries, survival probable	Severe life threatening injuries, survival uncertain

Table 1.1: Severity Levels

E0	E1	E2	E3	E4
Incredible	Very Low Probability	Low Probability	Medium Probability	High Probability

Table 1.2: Exposure Levels

C0	C1	C2	C3
Controllable in General	Simply Controllable	Normally Controllable	Difficult to control or Uncontrollable

Table 1.3: Controllability Levels

The standard further defines 4 ASILs (A - D) with A being the lowest and D being the highest levels of integrity. ASILs indicate the required safety measures that need to be

incorporated into the item or system. These ASILs are obtained as shown in equation 1.1 below.

$$ASIL = Severity(S) \times Exposure(E) \times Controllability(C) \tag{1.1}$$

Where each combination of S, E & C correspond to an ASIL level as shown in Figure 1.4. It also defines a QM level where no safety measures need to be incorporated and the standard development process is sufficient.

		C1	C2	C3
S1	E1	QM	QM	QM
	E2	QM	QM	QM
	E3	QM	QM	ASIL A
	E4	QM	ASIL A	ASIL B
S2	E1	QM	QM	QM
	E2	QM	QM	ASIL A
	E3	QM	ASIL A	ASIL B
	E4	ASIL A	ASIL B	ASIL C
S3	E1	QM	QM	ASIL A
	E2	QM	ASIL A	ASIL B
	E3	ASIL A	ASIL B	ASIL C
	E4	ASIL B	ASIL C	ASIL D

Figure 1.4: Levels of ASIL as defined in ISO 26262

We derive safety goals from the impact analysis and situation analysis as part of the HARA. The safety analysis correspond to inductive and deductive analysis. The safety goals are derived according to clauses 7.4.4.3 to 7.4.4.6 in part 3 of ISO 26262. The following holds true for safety goals:

- ASILs determined for a hazardous event are assigned to safety goals.
- Similar safety goals can be combined into a single safety goal.
- If a safety goal is satisfied by achieving a safe state or a set of safe states, the set of safe states are specified according to clause 8 of part 3.

- The safety goals and their ASILs are specified according to ISO 26262 part 8 clause 6.

The functional safety concept is used to derive the functional safety requirements from the safety goals and allocate them to the architectural elements and the subsystem as safety mechanisms. The functional safety concept defined in clause 8 of part 3 in ISO 26262 addresses the following:

- Fault detection and mitigation of failure.
- Safe states of operation and the transition of the system to those safe states.
- Fault tolerance mechanisms that mitigate the effect of a fault.
- Arbitration logic for choosing the appropriate control request from a set of control requests.

The functional safety concept is used to derive the functional safety requirements. These safety requirements are specified by taking into account the following:

- The system's operating modes.
- The fault tolerant time interval.
- The safe states of operation.
- The fault tolerant mechanisms.
- The emergency operation and the emergency operation interval in case the safe states cannot be achieved within the acceptable time interval.
- Warning mechanisms and function degradation concepts associated with a fault tolerant mechanism.

- The assumptions of driver related actions and the assumptions about the architecture that enable the satisfiability of the safety goals.

The functional safety requirements are derived for the hardware and software components of the architecture and allocated. We then derive the technical safety concept which specifies the technical implementation details for the safety mechanisms. At each stage of derivation of safety goals and their requirements, is an associated validation and verification reporting requirement.

1.3 Functional Architecture of Autonomous Vehicles

A functional architecture describes the interactions between the different functions and functional components of the system. The generic functional architecture for AVs has been explored previously by several authors [3], [30], and [34]. The architecture consists of the three components, 1) Sensing & Perception, 2) Decision & Control unit, and 3) Actuation & Stability unit. Each of the components is responsible for a different activity of the vehicle. The sensing and perception unit consists of sensors such as GPS, RADARs, LiDARs, and cameras as well as high-performance graphics processors and general processors. It is responsible for monitoring the environment, monitoring the state of the AV, and localization. The decision & control unit is responsible for planning the route of the vehicle, determining the next maneuver of the vehicle, management of the vehicle resources such as energy, and monitoring the general subsystem health. It consists of different elements, mostly high-performance compute units, which may be centralized or distributed across the vehicle. The stability & actuation component, consisting of low-level microcontroller units and actuators, is responsible for the traditional lateral and longitudinal control of the vehicle. All of the sensors, compute units and actuators are networked using a highly-reliable Ethernet and/or

CAN-FD backbone.

1.4 Reliability Engineering

The E/E architecture of an AV is expected to be resilient, and fault tolerant. Reliability and safety of an AV are correlated. Reliability focuses on the robustness of the system, and how it performs in the presence of faults. A fault is an abnormal condition, or defect, in a system that may result in the inability to perform its intended function. A fault can result in an error. Errors can result in a state of failure leading to a loss in quality of performance or system functionality. A fault at the component level is propagated at a function level and system level resulting in an erroneous state [17], [2] as shown in Figure 1.5. Faults are generally classified as systematic faults, which are inherent to the system design, and random faults, that occur due to probabilistic factors [13]. Let us define a few terms for system behavior on failure from reliability engineering:

Fail-operational (FO): The system can tolerate one failure and redundancy of resources allows the system to function correctly without human intervention. The system reaches a safe state of operation after a component fails.

Fail-safe (FS): By design, the system reaches a safe state of operation after one or several failures by default (passive fail-safe) or actively through some action taken by the system (active fail-safe).

Fail Silent (F-Sil): The component exhibits silent behavior and does not affect the system in any manner after one or more failures.

A system is said to be fault-tolerant if it can quickly reconfigure itself to continue operating with a minimal loss in performance after identifying a fault [4]. Safety mechanisms minimize

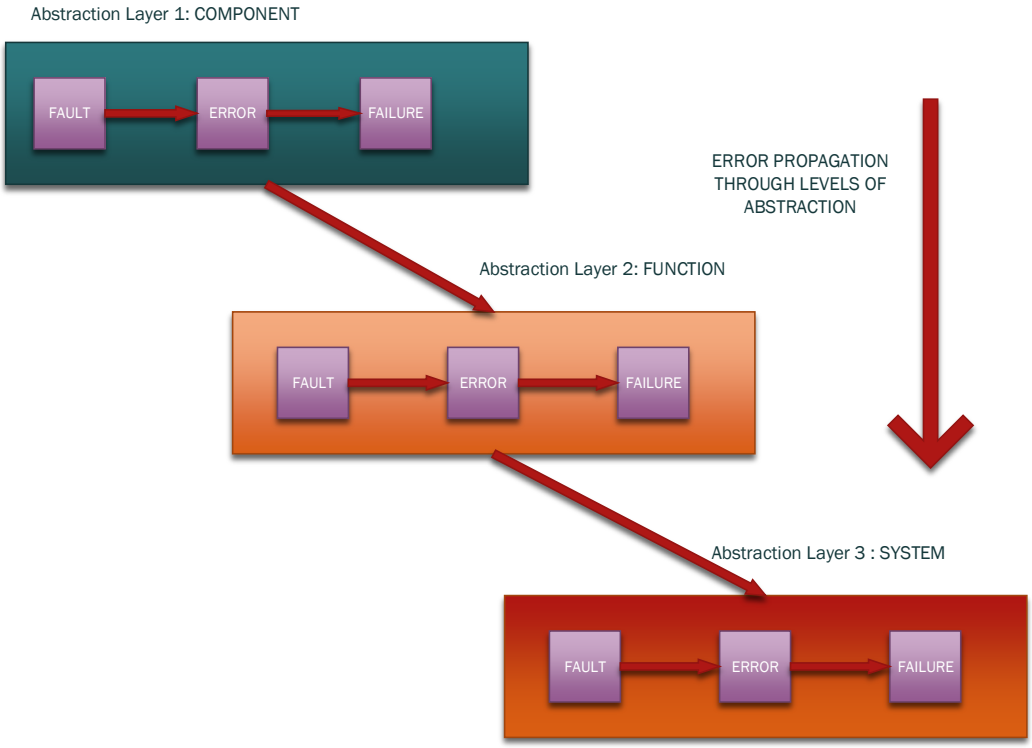

Figure 1.5: Fault propagation

the risk of a hazard by handling the faults at different abstraction levels. Implementations of fault-tolerance strategies are based on design requirements to mitigate faulty system behaviors. For example to avoid the babbling idiot problem we may implement a bus guardian to shut down the faulty electronic control unit which was flooding the bus with incorrect messages. Faults can be further classified as transient faults and permanent faults or as computation faults and communication faults at the system level. In the current work, we expect the system to be susceptible to the following types of faults as shown in Figure 1.6. These faults are examples of the generic malfunctions that any safety-critical system would suffer from and the listed faults are neither complete nor comprehensive.

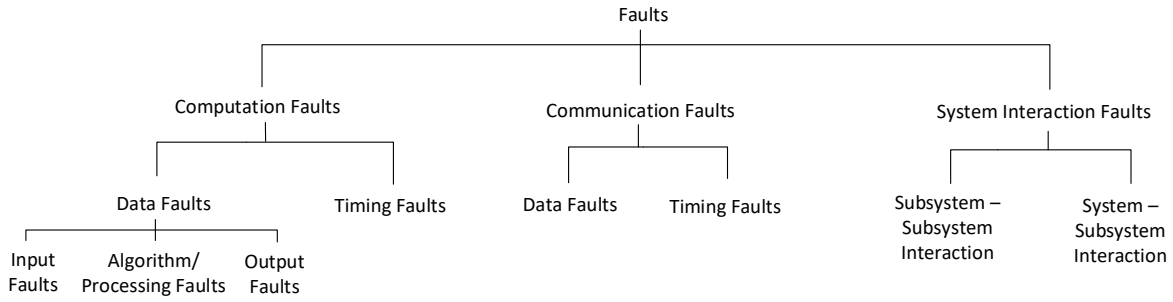

Figure 1.6: A short review of the possible malfunctions for Autonomous Vehicle

1.5 Problem Statement

The objective is to revise a part of the ISO 26262 process and develop a new approach to derive safety goals for the AV's development. The challenge in defining a new approach is in quantifying the decision-making capability of the vehicle and how this capability relates to functional safety. We also need to redefine ASIL and replace controllability with a realistic metric for assessing safety requirements in AVs.

1.6 Contributions

We propose a new approach to assess functional safety. The new method modifies the existing method with the addition of two extra steps, 1) Risk Mitigation Factor assignment and, 2) Influence Diagram assessment and defines AVIL - Autonomous Vehicle Integrity Level as shown in Figure 1.7. The Risk Mitigation Factor replaces controllability and the influence diagram assessment enables the designer to assess the AV's decision-making capability.

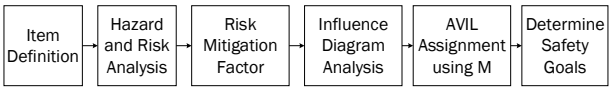

Figure 1.7: Summary of proposed practice

Our contributions in the current work can be summarized as follows:

- We define a new metric, Risk Mitigation Factor (M) to replace controllability used in the ASIL definition.
- We present a method to evaluate the functional safety and derive the safety goals for autonomous vehicles using the new metric.
- We present case studies highlighting the effectiveness of the proposed approach when compared to existing methods and techniques.
- We utilize Influence Diagrams to model the AV's decision-making capability and its expected behavior in a hazardous situation.

1.7 Related work

The standard ISO 26262 [11] is adopted from IEC 61508 [10] for the particular needs of the electrical/electronic systems within vehicles. Functional safety has been defined by ISO26262 as "The absence of unreasonable risk due to hazards caused by the malfunctioning behavior of E/E systems". The standard introduces ASILs or Automotive Safety Integrity Levels used in the development of the vehicle. ISO 26262 prescribes that safety is to be incorporated from the beginning of the inception phase and managed throughout the lifetime of the vehicle. The standard's major drawback being the reliance on the driver as the final form of redundancy in the system, which is no longer applicable to fully autonomous vehicles (SAE Level 4 and above).

System safety monitors that continuously monitor safety for autonomous vehicles have been proposed in [5] and [9]. The use of safety monitors, local and global, for different functions

[9], including the human-machine interface, autonomous driving monitor and driver monitor [5] have been advocated. The design of safety monitors themselves need to be guided by a set of rules or safety goals, and our approach specifies these safety goals. We also assume that these safety monitors are capable of assessing the risk of a driving situation continuously in real time as part of the decision & control component of the functional architecture.

The expansion of severity in the ASIL definition to a new level- ASIL E, covering multiple injury situations in cooperative platoons has been introduced in [20]. The work focuses on expanding the severity level to add an extra ASIL level while maintaining controllability as C3. The approach is The modification of how hazards are determined facilitates the evaluation of safety goals in autonomous systems. FMEA [16], FTA [6] or HAZOP [21] are helpful in assessing the system for safety, but they may not be sufficient to evaluate functional safety for autonomous systems. These methods do not account for the decision making capability of the system. The authors in the work [33] have shown that goal state notation (GSN) provides a better mechanism to evaluate automated systems when used with FMEA. The philosophy behind GSN and influence diagram assessment proposed in our approach are similar, with one major advantage of influence diagrams being the evaluation of decision strategies to obtain an optimal strategy reducing the risk. This advantage allows for evaluating the decision-making subsystem of the AV and deriving the required safety goals for it.

Assessing the correctness in decision-making capabilities of the system is part of evaluating it for the functional and the operational safety of the system. STPA or System theoretical Process Analysis [14] could be an alternative. STPA is found to be a better method for recognizing safety constraints of the overall system design while considering operational faults, as shown for the collision avoidance system in [28]. Other works have shown how STPA can be incorporated into ISO 26262 process to enable functional safety assessment of autonomous

vehicles [1], [29]. The combination of ISO 26262 and STPA has various advantages, as they cover both operational safety (safety concerning roadworthiness) and functional safety of the vehicle. ISO 26262 focuses on assessing the possible hazard when the system experiences malfunctions in that particular operational design domain. STPA focuses on inadequate control caused by human error, system interaction failures, and component failures. The works presented in [27] and [26] illustrate the application of STPA in a compatible manner to ISO 26262. A case study on functional safety analysis and safety goal evaluation is presented on an actuation control loop for an experimental autonomous vehicle. We apply our method presented in this work to the case study and compare the safety goals. The combined approach is still inadequate to assess the AV's decision-making capability since we are reliant on controllability and assign it to be C3 while assuming that all subsystems in an AV can be modeled as control-loops in a dynamic environment.

The FUSE project presented in [12] and the work preceding it in [31] describe iterative methods to perform hazard and risk analysis before the allocation of safety goals. The work presented in [31] uses an iterative method to define an item (an array of systems implementing a function [11]) and its attributes before starting the safety goal assessment process as per ISO 26262 and through function refinement. The FUSE project [12] uses a preliminary feature description where the function with the preliminary safety goal is refined iteratively with controllability being constant – C3. Our proposed approach redefines the ability of the AV to provide some form of control by the vehicle to mitigate the potential hazard caused by a fault without declaring that all sub-systems are uncontrollable.

In the works presented in [22] and [15], the authors present a very comprehensive view of the challenges faced in developing functional safety, with an emphasis on safe states for autonomous functions, as a key requirement for evaluating functional safety. A concept of assessing safety requirements by using prior risk and current severity for ADAS systems has

been presented in [7]. The work in [18] provides a framework for assessing hazards and risks of autonomous vehicles, but by retaining ISO 26262 with "Controllability as a metric of evaluation, does not comprehensively address safety assessment dilemma of every system being uncontrollable.

We present a new method to evaluate ASIL allocation for autonomous vehicles. The focus is to retain the design flow and safety concepts offered by ISO 26262 and combine it with the safe state of operation requirements for an autonomous vehicle. We provide a way to measure the "Controllability" of the vehicle or "the ability of the vehicle to reach a safe state without human intervention." We do not provide a refinement to any item definition or safety analysis, rather introduce a metric that can be used for ASIL allocation for the different items that are automated resulting in an Autonomous Vehicle Integrity Level or AVIL. The focus is on determining the required development integrity level for the hardware and software such that the safety goals are not violated and to ensure that the process of determining the integrity level can be standardized.

1.8 Structure of the Report

In chapter 1, we discuss the motivation and the contributions of this work. We also introduce the standard ISO 26262, discuss the generic functional architecture of an AV and general terms associated with reliability engineering. In chapter 2, we introduce the method to assess safety, the risk mitigation factor and the influence diagram based assessment. In chapter 3, we present three case studies on the engine management subsystem, the actuation system, and the decision & control system. In chapter 4, we discuss the results from the case study and conclude the topic. We also highlight possible areas of future work. In chapter 5, we summarize the contributions of the report.

Chapter 2

Proposed Factor and Method

In the current chapter we introduce Risk Mitigation Factor in Section 2.1 as well as the quantification of risk in Section 2.2. Section 2.3 introduces influence diagram assessment in the context of autonomous driving. We present the proposed method FuSAV in Section 2.4.

2.1 Risk Mitigation Factor: A time based metric

We have discussed the need for a new metric in Section 1.1, but what would be the characteristics of a new metric that can replace controllability? Any new metric must be replicable in engineering and reliable. The associated characteristics of safety management need to be reflected. Incorporating an aspect of body dynamics, controllability by the system and timeliness is required. "Risk Mitigation Factor" denoted by "Mx" is defined as *the capability of the system to perform a set of actions and reach a safe state of operation within the fault reaction time to avoid a potential hazard.* There are two aspects to the defined metric, one relates to the timing requirement and the other relates to the possible reactions the system can incorporate to respond to a malfunction. Hence, reflecting these two aspects, we further define a time "Risk Mitigation Window $W(M)$" (also used as "Window" interchangeably) as *the range of the fault reaction times for a given risk mitigation factor.* We also define a "Reaction r" as *the corrective response by the system to mitigate a potential hazard arising*

Table 2.1: Risk Mitigation Factors and associated scope of Reactions

Risk Mitigation Factor	Scope of reactions (r)	Window Requirements for Reactions
M0	The reactions available to the system enables it to function without any degraded performance Increase in risk within acceptable bounds after the occurrence of a fault The system has to reach a safe state within the specified T0 window of time	T0 Window T0 > T1 > T2 > T3
M1	Possible degraded performance of the system after recovery from a fault Increase in risk above acceptable levels after the occurrence of a fault Reactions within the T1 window of time to decrease the risk and reach a safe state	T1 Window T0 > T1 > T2 > T3
M2	Degraded performance of system after recovery from a fault A considerable increase in risk after the occurrence of a fault Reactions act within the short T2 window of time to decrease the risk and reach a safe state	T2 Window T0 > T1 > T2 > T3
M3	Degraded performance, but the system remains fail-operational Very high risk of harm after the occurrence of a fault Reactions act within the shortest window of time T3 to decrease the risk and reach a safe state	T3 Window T0 > T1 > T2 > T3

from a fault. We define the following elements (denoted by Mx) required for the Risk Mitigation Factor as shown in Table 2.1. The risk mitigation factor depends on the estimated time taken for reactions by the system. The reactions are intended to substitute the actions of a driver in a potentially hazardous situation. The factor substitutes for the controllability offered by the vehicle's capability to assess the environment and take mission-appropriate decisions in real time. The FRT and FTTI are the timing constraints for any reaction to be chosen, executed and completed safely to assure system safety. The time deadlines for

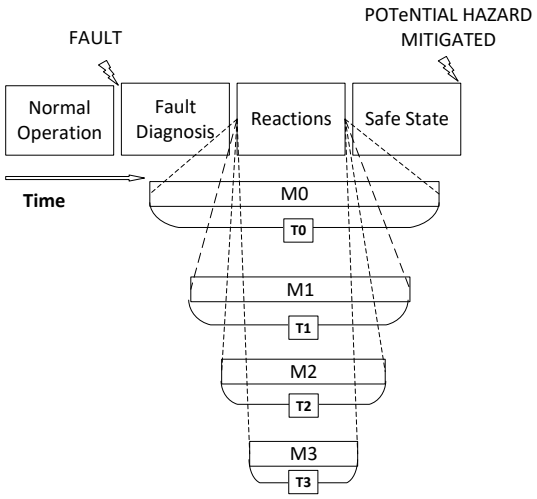

Figure 2.1: Risk Mitigation Factor, where T3 <T2 <T1 <T0

recovering to a safe state of operation after the AV experiences a malfunction are estimated by a designer based on statistical data or previous know-how. The risk mitigation window provides a conservative upper bound within which a safe state of operation needs to be assured before the risk becomes unacceptable. The different actions that the system can take in a hazardous situation to ensure safety depends on the reactions available to the system as well as the physical dynamics of the vehicle. The Risk Mitigation factor is defined for evaluating the plain item as is with any existing fault-tolerant mechanisms. Any item identified with a factor of M3 would be highly safety-critical to the functioning of the system. When such an item fails, without any functional safety mechanisms in place, the risk of a hazard

is very high. Figure 2.1 illustrates the concept of FRT based Risk Mitigation factor and the related windows. For every item, the window for the reaction can either be T0 or T1 or T2 or T3. The Risk Mitigation factor is defined accordingly.

By replacing controllability with the risk mitigation factor (M) we can retain the decomposition properties of ASIL assessment. The metric can be used to enumerate available reactions by the system to achieve a possible set of safe states. These actions are of the form of emergency maneuvers or passive/active safety mechanisms. Replacing controllability with the risk mitigation factor, retains the ASIL algebra, but the corresponding levels of development may not be the same. Finally, we shall define Autonomous Vehicle Integrity Level (AVIL) as shown in equation 2.1:

$$AVIL = S \times E \times M \tag{2.1}$$

The AVILs denote the requirement levels in development.

2.2 Quantification of Risk & Related Terms

To the existing framework of definitions described in Section 1.2 we make the following additions:

- Risk R = Probability of harm (P_H) x Severity (S) of harm
- Probability of harm $P_H = P_1 \times P_2$, where
  - P_1 is the probability of a hazardous situation occurring
  - P_2 is the probability of the hazardous situation leading to harm
- R' is the required residual risk

- $R''(r)$ is the expected residual risk after a set of reaction r are taken
- Fault Reaction Time (FRT) is denoted as, $FRT(f, a_f, s)$, where
  - $f : fault$ - the fault considered in analysis
  - $a_f : fault\_allocation$ - the subsystem in which the fault has occurred
  - $s : system\_reacting\_to\_fault$ - the system providing a reaction r_i to reach a safe state
- Similarly, FTTI is denoted as $FTTI(f, S)$
  - $f : fault$ - the fault considered in analysis
  - $S : System$ - the system considered in analysis

The definitions help us formulate the association between risk in different states, the fault reaction times and the reactions that the system can have for a given risk mitigation factor. By quantifying risk to be between 0 to 4, as a product of the probability of harm and severity, we provide the designer (and the AV) a method to assign a number to the risk R of a hazard due to a fault. This risk R has to be minimized to the required residual risk R‘ to assure a safe state of operation. Since severity of harm is a constant, the only possibilities to reduce the risk is to minimize either the probability of a hazardous situation occurring (P_1) or the possibility of that hazardous situation leading to harm (P_2). Any safety goal or derived functional safety concept can only attempt to minimize the probability of harm. The relationship between a risky state of operation after an event and the system entering a safe state of operation due to a reaction is modeled as an influence diagram described in the following Section.

2.3 Influence Diagram Assessment

We evaluate the effectiveness of reactions of the system by using influence diagrams. The reactions that satisfactorily reduce the risk and bring the AV to a safe state of operation are modeled as safety goals. Influence diagrams [25], [8] are decision graphs that represent all the factors affecting the decision making process for the given system. They are data-dependent cause-effect models with some uncertainty associated with some of the variables. An influence diagram N consists of a directed acyclic graph (DAG) $g = (V, \epsilon)$ representing the uncertainties, that are modeled as probabilistic events e, influencing a decision. These decisions by the system are modeled as reactions r. Deterministic events e_i are illustrated as ovals with a probability of $P(e_i) = 1$. The uncertain factors such as presence of an obstacle or the availability of a safety shoulder are modeled as conditional probabilities illustrated as ovals given by the conditional probability shown in equation 2.2 where A and B are assumed to be independent events. The probabilities of events ad decisions taking place simultaneously are assumed if empirical data is unavailable.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.2}$$

A reaction r_i, described in a rectangle, is taken if the preceding conditional probabilities are greater than a designated threshold i.e. reaction r_i is taken depending on the uncertain event e_i preceding it according to equation 2.2. Essentially we assure the system behavior in the event of a malfunction, or a certain behavior assures the presence of a malfunction for debug purposes when built accordingly. Finally, the value state or utility is represented using a diamond. The safe state modeled as a function of risk is the value state. The objective is to choose a set of reactions such that they minimize the risk modeled in the value state. Hence, an influence diagram re-purposed for functional safety in AVs as shown in Figure 2.2.

They can be represented as a four-tuple given by $N = (e, r, g, R''(r))$. Every reaction r_n has a measure of success - $MOS(r_n)$, and we define success as "Reaching a safe state when an independent reaction r_n completes within time $T(r_n)$". The risk in the safe state is assumed to be R'. Hence measure of success of reaction r_n is obtained as shown in equation 2.3:

$$MOS(r_n) = \begin{cases} R'/R''(r_n), & \text{if } R''(r_n) > R' \\ 1, & \text{otherwise} \end{cases} \tag{2.3}$$

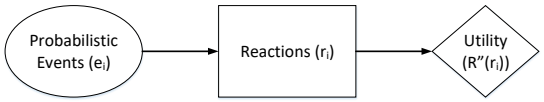

Figure 2.2: Influence Diagram Assessment

Influence diagrams support the solution of sequential decision-making with an optimal set of solutions when all the related variables are known. We assume that all the preceding information in a hazardous event is known and remembered by the system influencing the decision. The uncertain events e can be modeled either as the uncertainty in driving situations or the combination of failure rate of the component (λ_i) and Exposure (E) if known. Each reaction r_i has a direct impact on the probability of harm P_H by a factor k_i observed as a function of the reaction r_i and the current value of risk R by the system.

$$k_i = F(r_i, R)$$

By varying the different reactions r_i, and the expected k_i values, we can evaluate if the hazardous event has been mitigated and the risk has been reduced to a reasonable level.

Every reaction r_i is treated as an independent reaction to a fault. Let us assume that for an example component the failure rate is $\lambda_i = 0.2$, with an exposure of a given driving situation $E = E2 = 3$ or between 10% - 40% of the time for fault f. We get the probability of the

event $e_i = 0.6$ with a risk $R = 3.6(S = 4 \& P_H = 0.9)$. We now assess a system's reaction r_i to fault f and its capability to reduce the risk R to an acceptable level R'. Assuming that we have a reaction r_1 with $k_i = 0.2$. The probability of harm is reduced by a factor of k_i after a reaction completes successfully. The expected residual risk would be R"$(r_1) = 0.72$. If the expected residual risk R"(r_1) is lesser than the required residual risk R', the reaction can effectively deal with the fault and restore the system to a safe state. A set of reactions are found if an individual reaction r_i is insufficient, where each reaction is independent and has an impact on the risk R as shown in Figure 2.3.

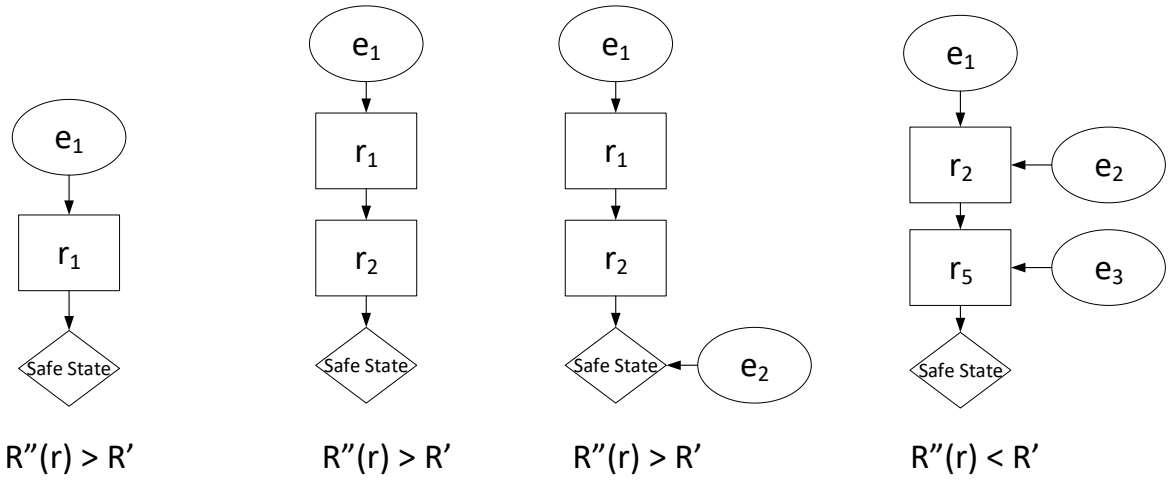

Figure 2.3: Determining Sufficient set of reaction r for given set of events e

For a given set of reactions r, we obtain the expected measure of success $MOS(r)$ using the weighted product of probability of a reaction r_i and its measure of success $MOS(r_i)$ shown in equation 2.4. The designer sets the MOS', which is the required measure of success for a set of reactions r. The expected measure of success MOS needs to be greater than the required measure of success MOS'.

$$MOS(r) = \sum_i (P(r_i/e_i) \times MOS(r_i)), \forall r_i \in r, \forall e_i \in e \tag{2.4}$$

2.3.1 Quality of Reactions

Reactions substitute the possible actions by a driver in an AV. These are essentially response to a fault, hence they may incorporate certain fault-tolerant behaviors. To better define a reaction the following rules apply.

- Reactions shall only reflect whatever is physically capable by a system. A good example would be *r: Send braking signal within x mS* whereas a reaction *r: Brake within x mS* for a vehicle traveling at high speeds may not be physically possible.

- Reactions correspond to one action by the vehicle. Compounded reactions such as *r: Send braking signal within x mS and halt within y mS* would be incorrect.

- The time T(r) for a reaction needs to be determined through simulation or experimentation. It has to account for the maximum time for the reaction to be initiated, executed, and completed as intended i.e., a reaction to change lanes should take at most T(r) time units to result in the vehicle switching lanes.

- The k_i value of a reaction is lower if the potential of the reaction to mitigate the fault is higher. The value ranges between 0.1 - 1. For reaction *r: Switch to secondary processor* in a duplex architecture the k_i can be assumed to be equal to 0.1 as it is highly effective and can immediately reduce the probability of harm.

2.3.2 Satisfiability Criterion For Reactions

A set of reactions $r = \{r_1, r_2, ...r_n\}$ are sufficient for the system to reach a safe state of operation after a fault f if the following satisfiability conditions are met:

- $R''(r_n) \leq R'$, i.e. where the residual risk $R''(r_n)$ after r_n completes is lesser than the required residual risk R'.

- $\sum_i (T(r_i)), \forall r_i \in r, \leq FRT(f, a_f, s)$, i.e. the sum of execution times of all reactions is lesser than or equal to the fault reaction time for the system.

- The computed $MOS(r)$ value needs to be greater than the required value MOS' for the set of reactions r to be sufficient.

The set of reactions obtained form a safety strategy that can be implemented in order to meet the required safety goals as part of the functional safety concept or can be framed as a safety goal. We note that the order of the reactions in the set may impact the safety of the vehicle. We focus on satisfiability in the current work and not on optimizing the set of reactions which is a potential future direction of research.

2.4 Method to Evaluate Safety

The AV is capable of making decisions regarding tactical and strategic operations, and these decisions include the response to a fault or a set of faults. Each reaction is a response to a fault, and a satisfactory set of reactions are incorporated in the design of the AV formulated as a safety goal. We make the following assumptions:

- Single Point Faults lead to hazards when propagated through the various layers of abstraction; an unmitigated fault leads to an error. Errors lead to system failure and a failure results in a potential hazard.

- The fault detection time is known for the entire system and the set of faults that are being analyzed.

- The time taken to choose reactions after detection/identification of a fault is negligible.
- The set of reactions possible by the system are independent of each other and may be executed sequentially or concurrently in the presence of a malfunction.

We present FuSAV - Functional Safety in Autonomous Vehicles, a method to evaluate an Item (interchangeably used with system) and determine the safety goals using risk mitigation factor M. We also identify the set of reactions r that satisfy the FRT requirements. The

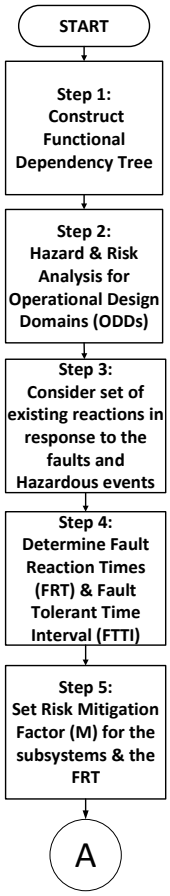

Figure 2.4: FuSAV: Functional Safety Assessment for Autonomous Vehicles Steps 1-5

FuSAV method is as shown in Figure 2.4 and Figure 2.5. Initially, we consider the given system (plain Item) and construct the functional dependency tree (shown in step 1 of Figure

2.4). The functional dependency tree illustrates the hierarchical control dependency and the data dependency between the top level system and its subsystems. It also defines the boundaries of the system as functions and their interactions. Feedback loops from sensors are assumed to be part of the function. Therefore any closed loop system is represented as a tree with a hierarchal relationship between systems.

Then in step 2 we the hazard and risk assessment (HARA) for different operational design domains(ODDs) [19] and possible malfunctions are analyzed as described in ISO 26262. There are two significant differences in this step as compared to the traditional HARA used in ISO 26262 - 1) Controllability is not used as a factor to evaluate the hazardous event and the driving situation. Only severity and exposure are used. 2) We calculate the risk of harm on a scale of (0-4] in a hazardous event. The severity of harm on an integer scale of [1-4] and the probability of harm on a scale of [0-1] is used to compute the risk. We list out all the potential malfunctions and hazardous events exhaustively in this step. We choose the fault to be analyzed in the consequent steps.

In step 3, we list the built-in reactions that are already available for the system as part of either the plain item or the legacy system. These existing reactions are considered as-is without any modifications or additions in response to the fault in the analysis. The system is capable of providing these reactions without any further development or additional functionality. In step 3, we also consider any potential reactions that can be provided by other legacy systems in the vehicle to mitigate the risk of a hazard.

Once a comprehensive list of reactions are generated, we estimate the FRT and FTTI for the system based on empirical vehicle data or designer experience for a particular driving situation and possible hazard in step 4. This FRT potentially can be propagated bottom-up or top-down on the functional tree obtained in step 1 depending on how much information is available about the system. We note that the FRT and FTTI rely on the body dynamics

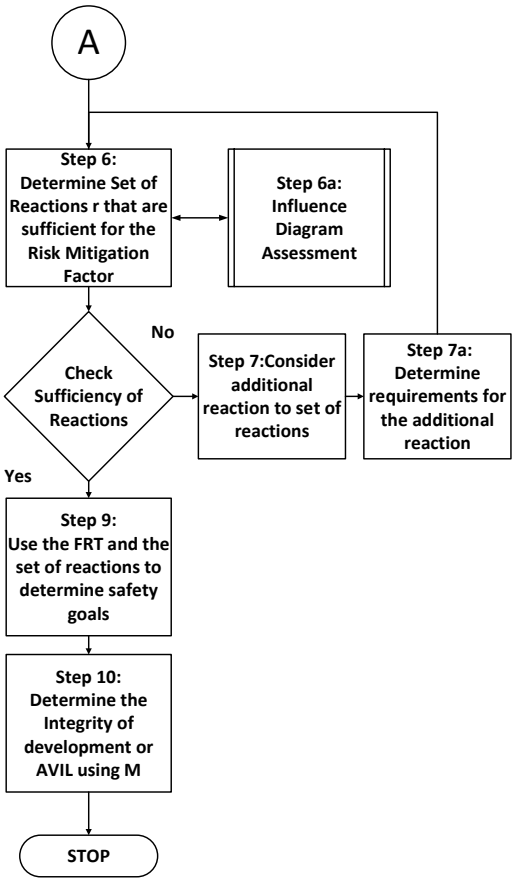

Figure 2.5: FuSAV: Functional Safety Assessment for Autonomous Vehicles Steps 6-10

of the vehicle (factors such as acceleration, mass, speed, the coefficient of friction between the tires and roads, etc.) and the driving scenario (urban roads, freeway, rain, etc.).

In step 5, we divide the expected fail-operation time of the vehicle into four windows of FRT for risk mitigation factor assessment. We set the corresponding risk mitigation factor M based on the FRT value.

In the iterative steps- 6 and 7 shown in Figure 2.5, We check for sufficiency of set of reactions r. If there is no set that satisfies the satisfiability conditions, then in step 7 and 7a, we add a new reaction to the system and define its functional characteristics. We reiterate steps

6-7 as part of step 8 until the set of reactions are satisfactory according to the criterion established previously in Section 2.3.2. Steps 6-8 allow for the addition of a reaction limited by the conditions established in Section 2.3.1 to reduce the probability of harm. We assume the joint probabilities of events and reactions if reliable data is not present.

In steps 9 and 10, as shown in Figure 2.5, we use the set of reactions to determine the safety goals and the AVIL level using the risk mitigation factor M. The safety goals are formulated in accordance to the guidelines prescribed in ISO 26262 part 3 clause 8.

Chapter 3

Case Study

In the current chapter, we evaluate AV subsystems using the base ISO 26262 approach in Section , the ISO 26262 with STPA approach as described in [1] in Section and the FuSAV approach presented in this paper in Section.

3.1 Autonomous Vehicle Architecture

A fully autonomous vehicle (SAE Level 5) operates without human input and intervention. It is capable of completing its mission regardless of the malfunctions. The AV is fail-operational for the entirety of its journey between two locations of interest. We assume that the underlying electronic/electrical architecture is fault-tolerant exhibiting fail-safe behavior. The functional architecture of an autonomous vehicle described in [1] is as shown in Figure 3.1 is similar to the described in Section 1.3. The autonomous vehicle has three main systems:

- The decision and control system (S0) responsible for trajectory planning, maneuver planning, and the driving strategy.
- The actuation and stability (S4) responsible for the steering system, braking system, and the engine system as well as their actuators.
- The sensing and perception system (S8), responsible for the underlying sensors, data fusion, environment modeling, and vehicle localization.

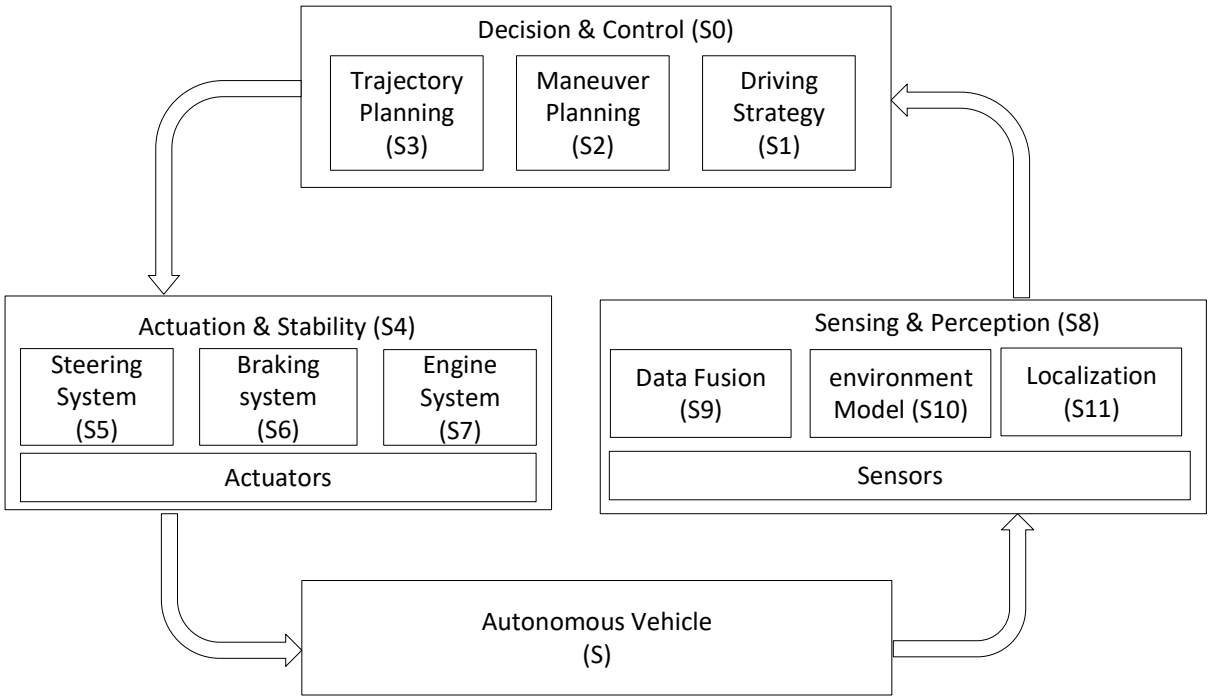

Figure 3.1: Functional Architecture of an AV described in [1]

The high-level functional architecture outlines the complex interactions between the different subsystems of the architecture. The decision and control systems are dependent on the sensing and perception systems to provide information about the environment as well as the state of the vehicle. The actuation and stability systems depend on the decision and control outputs. One can easily notice that any fault in this tightly coupled closed loop federated architecture can lead to a potential hazard. A simple delay in sensing or an error in computation can potentially lead to harm in an extremely otherwise normal driving scenario. Consider the example of an urban environment, where a potential obstacle is not detected in time, it could lead to a traffic incident where users and onlookers may be injured. Therefore, it becomes essential to capture as many aspects of the architecture as possible before beginning any safety evaluation. For such a complex system of systems, there are different types of faults possible at different levels of abstraction from the vehicle level to

the component level. It may be difficult to enumerate all possible malfunctions, but it is necessary that we exhaustively list out as many potentially hazardous driving scenarios.

3.2 Case Study: Engine Management System

In the following case study we evaluate an engine control subsystem using the base ISO 26262 approach, the ISO 26262 with STPA approach as described in [1] and the FuSAV approach presented in this paper. We consider the engine system - S7 (interchangeably used with engine management system), for the architecture presented in [23], whose primary function is to provide an appropriate acceleration by providing the engine torque and maintaining the pressure for an AV. The control unit reads the vehicle speed V, the engine crankshaft speed R, the transmission ratio TR, the digital equivalent of an acceleration pedal position G and P for engine pressure and T for the throttle as outputs with a feedback mechanism as shown in Figure 3.2. There is a feedback mechanism to read the throttle value and the pressure in the combustion chamber. There are electromechanical valves that are opened or closed depending on the signals from the logic processing unit. The throttle produced is a function of the four input signals.

$$T = f(G, V, R, TR)$$

3.2.1 ISO 26262

We apply the traditional ISO 26262 process by utilizing the HARA as is to derive the safety goals and allocate ASIL for the system. We find that according to Table 3.1 and the results from for the malfunction of accelerating beyond what is intended in a risky scenario leads to

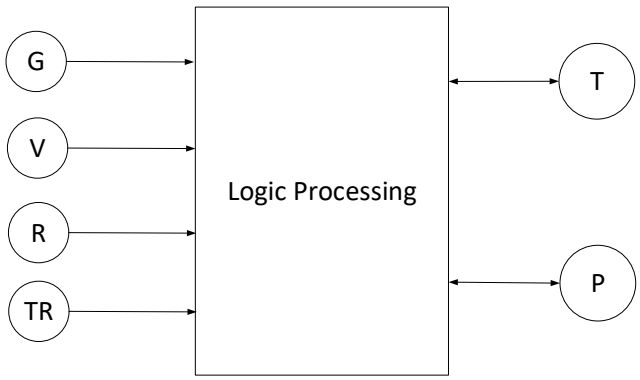

Figure 3.2: Example engine management system

an ASIL allocation of ASIL D for the hazardous event *HE.1: AV accelerates beyond acceptable limits in an urban environment resulting in possible collision with other traffic or pedestrians.* We refer to the work in [23] to obtain a detailed set of safety goals. The exposure being E3 i.e., driving in urban areas is a very common driving scenario, severity being S3 i.e., there is potential of serious injuries to pedestrians using the road and controllability being C3 or uncontrollable, since for the AV the users cannot provide an action to mitigate the hazard. Since, we do not have the driving data for determining S and E values accurately, we have chosen E3 and S3 by considering a conservative bound for illustration. For the autonomous vehicle with safety goals as shown in Table 3.2, the required ASIL for development would be ASIL D, which is the highest level of integrity in development. The time bounds t1 and t2 shown in the Table 3.2 are derived from the vehicle body dynamics. The time bound t1 is a strict upper bound beyond which a torque buildup in the engine leads to a potentially hazardous situation in an urban scenario. To prevent the vehicle from *HE.1*, an ASIL D in the development integrity is required. The safety goal SG2 does not need an ASIL D integrity requirement as an unintended decrease in engine torque at most would lead to a rear-end collision and is less safety-critical as compared to SG1 requirement (in terms of ASILs). Since, SG1 is a stricter requirement, the overall development of the engine

management system requires an ASIL D in development. We also note that the SG5 relating to sensor integrity is also an ASIL D requirement. The functional safety concept and technical safety concept are derived with an emphasis on iterating scenarios where the safety goal can potentially be violated according to ISO 26262 part 3 clause 8 described in Section 1.2.

Table 3.1: Selective Hazard and Risk Analysis Results

Item	Malfunction	Driving Situation	Risk Scenario	Hazard
Engine System (S7)	f1: Accelerates more than intended	Urban driving environments with medium speeds between 25 mph and 40 mph	Traffic enroute or presence of obstacles	Collision possible with infrastructure or pedestrians
	f2: Missing feedback from sensors regarding vehicle state	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 300m$	Collision possible with traffic or infrastructure
	f3: Erroneous throttle valve signal	Freeway at high speeds (> 40 mph)	Minimum preferred distance to obstacle exceeded	Collision possible with traffic or infrastructure

Table 3.2: Safety Goals For the System Derived from HARA for Fault f1 - ISO 26262

Safety Goal
SG1: Avoid unintended increase in engine torque in Urban areas beyond t1 time period
SG2: Avoid unintended decrease in engine torque in Urban areas beyond t2 time period
SG3: Avoid unintended acceleration in urban areas beyond t1 time period
SG4: Avoid unintended deceleration beyond t2 time period
SG5: Ensure integrity in sensor feedback

3.2.2 ISO 26262 & STPA

We apply STPA with ISO 26262, we follow the procedure described in [1]. The procedure can be summarized as follows:

- Apply STPA step 0
- Apply ISO 26262 HAZOP
- Apply STPA step 1
- Apply STPA step 2

The step 0 of the STPA is used to define the item, identify system level hazards and accidents, construct the control loop at a given abstraction level, and identify the safety constraints. At the system level we identify possible accidents such as *AC.1: The AV collides with traffic in urban roads* due to the hazard *HA.1: The AV lost control due to unintended dangerous acceleration* which corresponds to the hazardous event described previously. The safety constraint *SC.1: The AV must not have missing sensing feedback values* is the high level system safety constraint. The difference between a safety goal and a safety constraint is that the safety goal is a requirement that needs to be satisfied whereas the safety constraint is a system level constraint that denotes a limitation associated with the safety requirement. The high-level abstraction of the engine management system is as shown in Figure 3.3. The control structure is used to identify the boundaries of the item and aids in defining the item for the HARA process. The system level accidents and the list of hazards are taken as inputs to HARA process. Then the list of operational situations such as *OS.1: driving in urban environment.*

We apply the HARA process with the step 0 outputs as inputs as shown in Table 3.3 . We derive similar safety goals (Table 3.5) as earlier (Table 3.2) and find that the ASIL level is ASIL D. This is again due to controllability being considered as C3. Since we are considering the same hazardous event, we assign severity to be S3 and the exposure to be E3. In step 1 we identify the higher level safety control actions such as *CA.1: appropriately opening the throttle valve* shown in Table 3.4 and iterate the resulting unsafe control actions possible by

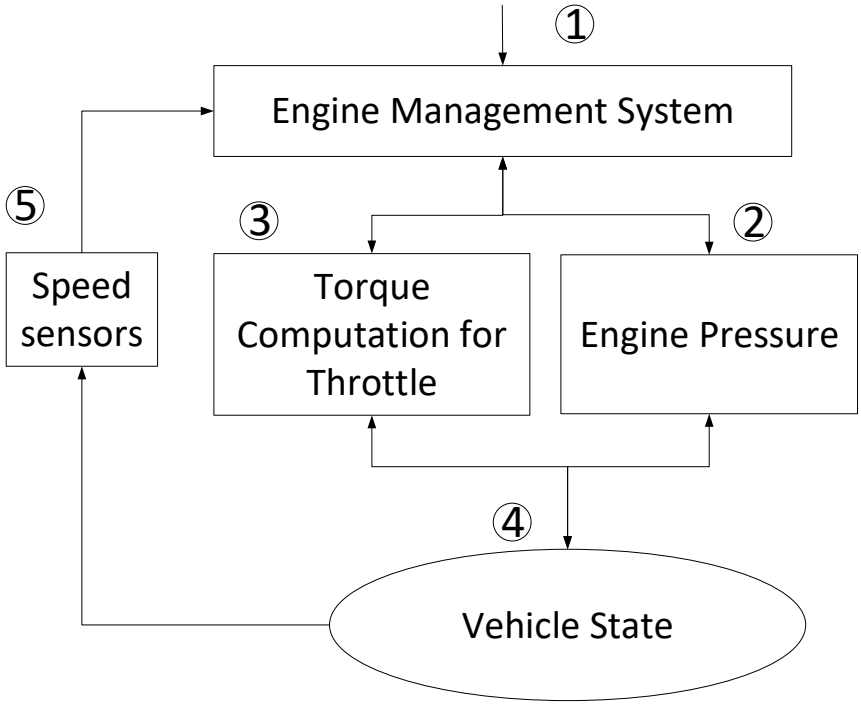

Figure 3.3: Control Loop for Engine Management System.

the system by evaluating four hazardous types described in [14], i.e., missing or incorrect control action, incorrect timing of control action, incorrect sequence of control actions and safe control action provided for too long or too short time period. In step 2 we perform the causal factor analysis as shown in Figure 3.4 to derive the causal scenario in which an unsafe control action results in an accident for an AV. We get the following unsafe control actions as shown in Table 3.6 for the control actions CA.1 and CA.2 specified in 3.4. *UCA.1: The throttle valve is closed too late* results in the new safety constraint that the *AV must provide the signal to close the valve with appropriate timing.* The step 2 results are used to build the safety concept. These high-level safety constraints can then be used to identify the technical safety requirements as well as the functional safety requirements. We ask the interested reader to refer to the following works presented in [1] to gain a better understanding of how this is achieved.

Table 3.3: Selective Hazard and Risk Analysis Results for Engine System

Item	Malfunction	Driving Situation	Risk Scenario	Hazard
Engine System (S7)	f1: Accelerates more than intended	Urban driving environments with medium speeds between 25 mph and 40 mph	Traffic en-route or presence of obstacles	Collision possible with infrastructure or pedestrians
	f2: Missing feedback from sensors regarding vehicle state	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 300m$	Collision possible with traffic or infrastructure
	f3: Erroneous throttle valve signal	Freeway at high speeds (> 40 mph)	Minimum preferred distance to obstacle exceeded	Collision possible with traffic or infrastructure
	f4: Pressure build-up in engine due to inoperable valve	Freeway at high speeds (> 40 mph)	Obstacle present	Collision possible with traffic or infrastructure

Table 3.4: Example STPA Step 1: Safe Control Actions

System	Safe Control Action
①	CA.1: Appropriately opening the throttle valve
①	CA.2: Appropriately closing the throttle valve
②	CA.3: Calculate required pressure value accurately
③	CA.4: Calculate required torque value accurately
③	CA.5: Change torque value

Table 3.5: Safety Goals For the System Derived from HARA for Fault f1 - using ISO 26262 with STPA

Safety Goal
SSG1: System MUST avoid unintended increase in engine torque in urban areas
SSG2: System MUST avoid unintended decrease in engine torque in Urban areas beyond t2 time period
SSG3: System MUST avoid unintended acceleration in urban areas beyond t1 time period
SSG4: System MUST avoid unintended deceleration beyond t2 time period
SSG5: System MUST ensure correct and timely sensor feedback
SSG6: System MUST ensure correct sequence of control actions

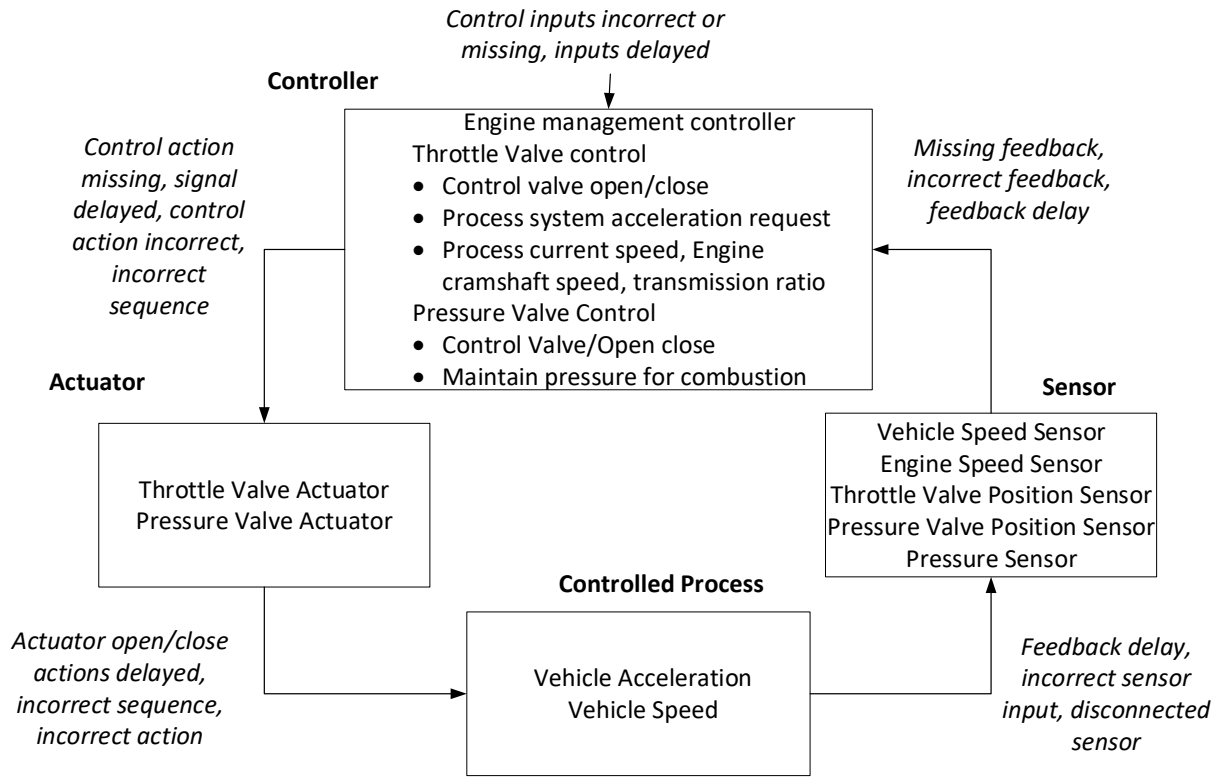

Figure 3.4: Causal Factors Analysis for Control Action CA.1

Table 3.6: Example STPA Step 2: Unsafe Control Actions

System	Safe Control Action
①	UCA.1: The throttle valve is closed too late
①	UCA.2: The throttle valve is opened too late
①	UCA.3: The throttle valve is opened too early
①	UCA.4: The throttle valve is closed too early

3.2.3 FuSAV: Assessing Functional Safety

We apply the FuSAV method on the engine management system shown in Figure 3.2. In step 1, we derive the functional dependency tree showing the control dependency in a hierarchical manner when traversed from top-down and the data dependency of the top-level system on the lower subsystem when traversed bottom-up shown in Figure 3.5. We then perform

HARA in step 2 for the functional dependency tree (obtained in step 1), which results in the same table as the ISO 26262 process (Table 3.1). For each of the HEs we only consider severity and exposure, as explained in previous Sections controllability is not used. In step 2, considering the severity as 4 and probability of harm as 0.7, the risk R is 2.8. Assuming that the designer expects the required risk R' to be 0.8 in the safe state (acceptable risk), We enumerate the available reactions to system as a whole (considering the entire vehicle as well as the system in study) shown in Table 3.7 as well as the probabilistic events that may affect the reactions shown in Table 3.8.

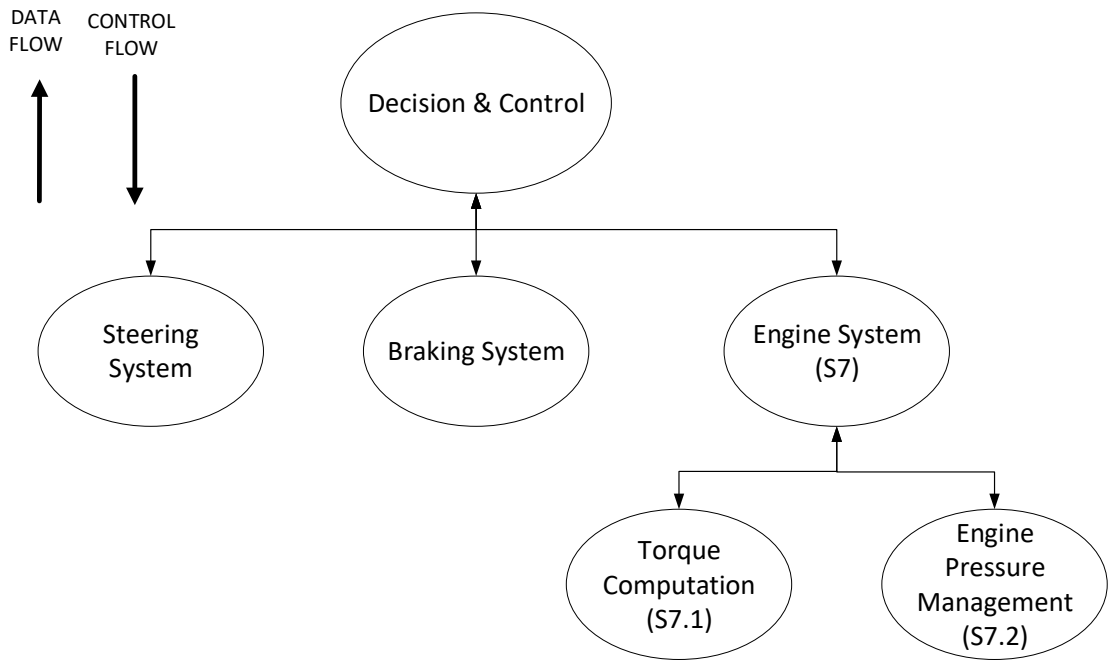

Figure 3.5: FuSAV Step 1: Functional Dependency Tree

We determine the FRT and the FTTI values based on the vehicle body dynamics in step 4 and set the values for each subsystem in step 5 as shown in Figure 3.6. The risk mitigation

Table 3.7: Step 3: Reactions for fault $f1$

Reactions r_i	$T(r_i)$ mS	K_i
r_1: Decrease engine pressure to offset thrust	120	0.5
r_2: Compute new throttle outputs every 50mS	50	0.5
r_3: Apply braking to reduce acceleration	50	0.3
r_4: Halt completely	100	0.2
r_5: Decrease vehicle speed through change in transmission	50	0.5
r_6: Change lane	100	0.4

Table 3.8: Events affecting Reactions for fault $f1$

Events e_i
e_1: Urban area with possible pedestrian activity
e_2: Obstacles detected with system in relatively high speed

factor for the fault in analysis is determined in this step and assigned depending on the FRT range. The fault detection time (FDT) plays a very important role in estimating FRT. We assume that the FDT is uniform for all faults and is known beforehand. The FDT is assumed to be 100 mS. Here, we estimate the FTTI, FRT and FDT values based on assumed behavior and response for the case study, in reality these are determined by rigorous experimentation and statistical inference of the vehicle body dynamics.

The total fail operation time for the system is assumed to be 2 seconds and the window for the FRT's are divided in four equal intervals to obtain four windows as shown in Table 3.22. We suggest that the best method to divide the window is based on statistical inference and designer experience. For now, we assume that the windows are as shown in Table 3.22. The FTTI of the system is assumed to be 500 mS. Based on the FRT intervals, the system is assigned a risk mitigation factor of M2 as the FRT(f1, S7, S7) = 150 mS for fault in analysis. For our current example the FRT values have been propagated top-down as shown in Figure. FRT propagation across the functional dependency tree can be done top-down or bottom-up based on the designer's preference and the availability of information about the system.

Table 3.9: Step 5: Risk Mitigation Factor Windows

Risk Mitigation Factor Mx	Window in mS W(Mx)
M3	0 - 125
M2	126 - 250
M1	251 - 375
M0	376 - 500

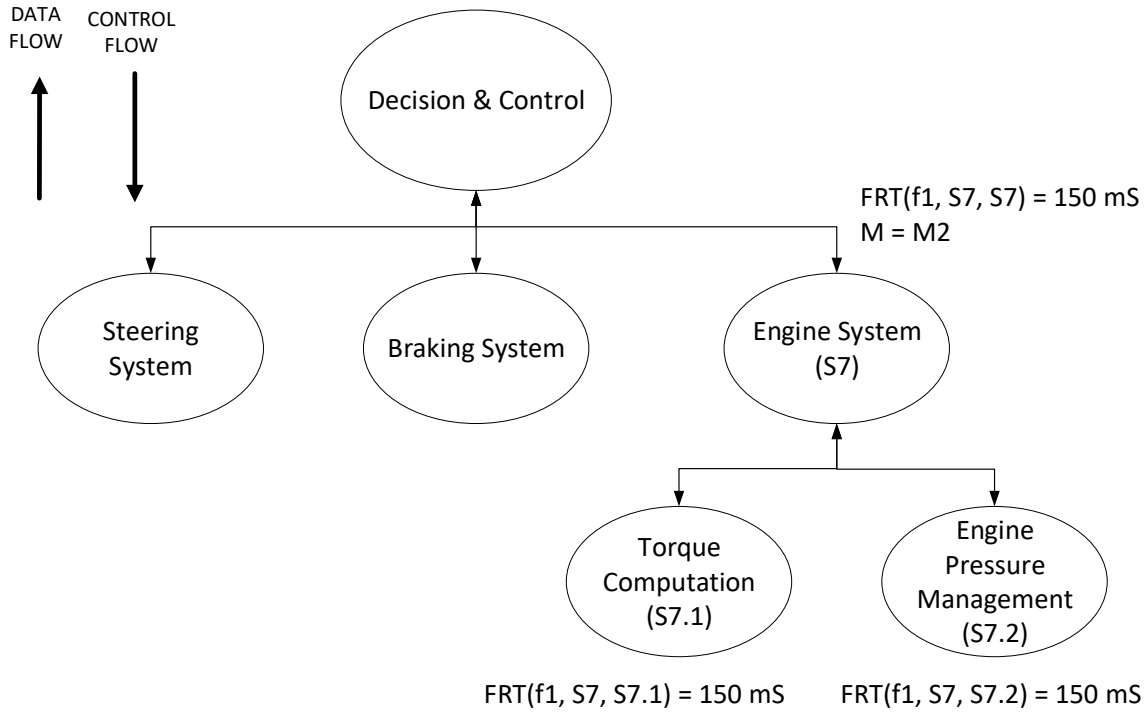

Figure 3.6: FuSAV Step 4& 5: Functional Dependency Tree with FRT and Risk Mitigation Factor Assignment

In step 6, we determine the sufficient set of reactions r, that need to be incorporated into design by utilizing ID assessment and the measure of success $MOS(r)$ metric as shown in Figure 3.7. The joint probabilities of $P(r_1 \cap e_1)$ is assumed to be 0.03. The probability of event e_1=0.1, which is a combination of the rate of failure of the engine management system

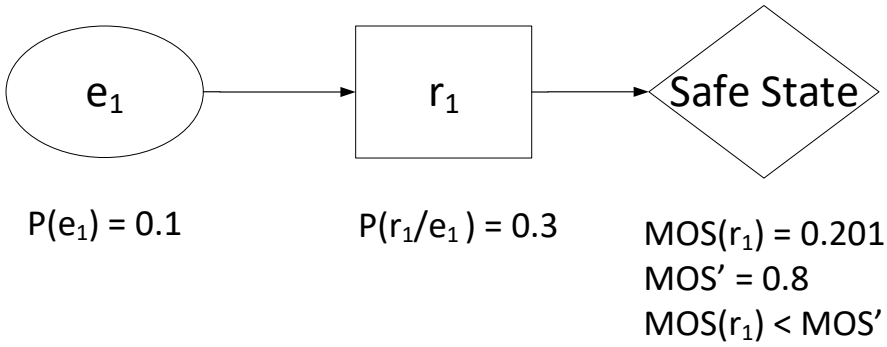

Figure 3.7: Step 6-6a: Influence diagrams for functional safety assessment

and the driving scenario described in Table 3.8. We find that the reaction r_1 is insufficient in reducing the risk, $R''(r_1) = k_1 \times R = 1.4 > R'$. Since r_1 by itself is insufficient, a set of reactions r are computed. The set of reactions need to satisfy the satisfiability criterion. Steps 6 & 7 are iterated until a sufficient set of reactions r are found. It does not reduce the risk satisfactorily in the given situation. The Hence, we reiterate steps 6-7 to obtain a satisfiable set of reactions. We assume the joint event probabilities required to compute the conditional probability.

The designer sets the MOS' value = 0.8 as the threshold for the set of reactions r to ensure that the system satisfactorily reaches a safe state of operation. We obtain a set of reactions $r = \{r_2, r_3, r_5\}$ as shown in Figure 3.8 that satisfy our criterion. The reactions currently shown and described are sufficient as finding the optimum set of reactions is not the scope of this paper. In step 9 as shown in Table 3.10, we specify the set of safety goals using the template shown in Figure 3.9. These safety goals need to be formulated in accordance to the clause 6 of part 8 specified in ISO 26262, but for now we use a template with a guide word MUST as an example safety goal constructor to obtain safety goals comparable to the ISO 26262 with STPA method.

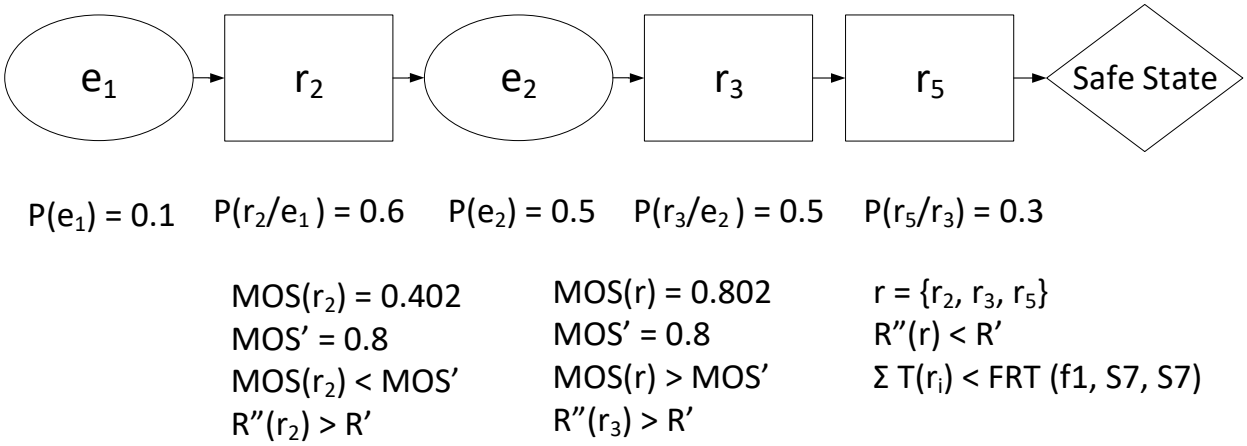

Figure 3.8: Step 7: Additional reactions added to obtain a set of reactions satisfying the satisfiability criterion

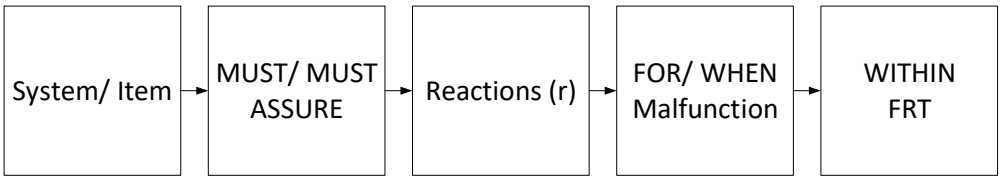

Figure 3.9: Step 9: Safety Goals template

The reactions are enforced as safety goals and the functional safety requirements during the design process. The functional safety concept and the technical safety concept are derived in accordance to the prescribed methodology in ISO 26262. The advantage of FuSAV is clearly seen in analyzing fault f1. The ASIL is reduced from ASIL D to AVIL C (for now we assume equivalence of ASIL and AVIL, i.e., ASIL C is equivalent to AVIL C). There is a very low chance that a system's safety requirement is overdesigned. This is because the risk mitigation factor associates the system's capability to take evasive actions or initiate safety mechanisms in the presence of a malfunction to mitigate the risk of a hazard. Furthermore, the designer and safety engineer can decide on implementing a set of reactions that are optimal in mitigating the potential hazard. The decision-making module and its subsystems

Table 3.10: Safety Goals For the System Derived from HARA-FuSAV for Fault f1

Safety Goal
FSG1: System S7 MUST avoid unintended increase in engine torque in urban areas within t1 period of time
FSG2: System S7 MUST avoid unintended decrease in engine torque in Urban areas beyond t2 time period
FSG3: System MUST avoid unintended acceleration in urban areas beyond t1 time period
FSG4: System MUST avoid unintended deceleration beyond t2 time period
FSG5: System MUST ensure correct and timely sensor feedback within FRT(f1, S7,S7) = 150 mS
FSG6: System S4 MUST ensure braking to reduce the unintended acceleration in urban roads within T(r_3) = 20 mS
FSG7: System S7 MUST reduce unintended acceleration by reducing the transmission value TR in urban roads within T(r_5) = 50 mS
FSG8: System S7 MUST compute thrust values every T(r_2) = 50 mS

can be designed according to the derived safety goals. The risk mitigation factor denotes the FRT requirements whereas the reactions and the ID assessment help model the intelligence and decision-making capability of the system and how this capability can be modeled as the system's equivalent of controllability. An optimal set of reactions denote a risk mitigation strategy that can be taken by the system in the presence of fault.

The different methods explored in the case study result in similar safety goals and functional safety concept as it can be seen in the Tables 3.2, 3.5 and 3.10. We notice that the goals derived from ISO 26262 & STPA in Table 3.5 are much more stringent than the ones in Table 3.2 due to the keywords MUST and SHALL prescribed in STPA. It may be very difficult or expensive to develop functions which will not fail. The safety goals in Table 3.5 are more comprehensive as they include the safety goals in Table 3.2 as well as add newer goals for the correct sequence of control actions by the loop. We notice that the safety goals derived from FuSAV in Table 3.10 go one step further and define the reactions that other systems can provide to assure system safety. The results in Table 3.10 show that a risk mitigation strategy can be adapted by the system to deal with the malfunctions and hazardous events.

The safety goals derived from FuSAV correspond to a wide variety of driving situations and provide the system requirements to resolve risky driving scenarios.

3.3 Case Study: Actuation System

The case study compares results with the work in [26] where the authors have applied STPA for an automated actuator control system. We apply the FuSAV method on the control structure of the actuation system described in [27] as shown in Figure 3.10. In Step 1 we

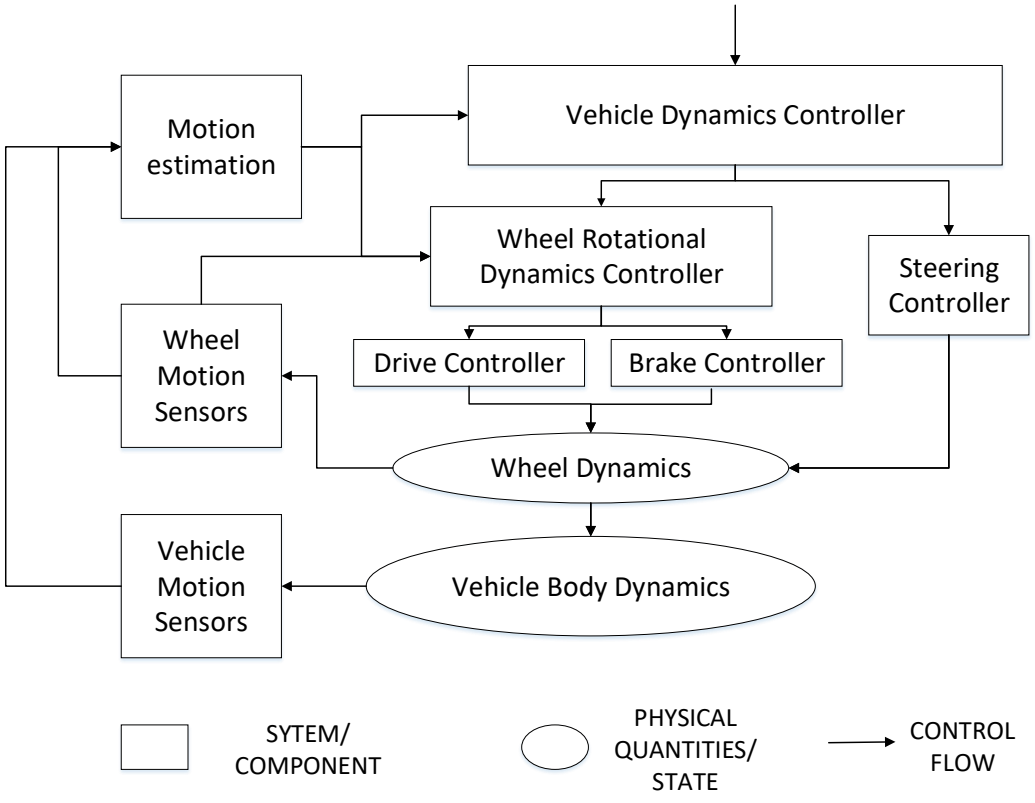

Figure 3.10: Control system of the actuation system in study [27]

construct the functional dependency tree to obtain a hierarchical relationship between the different subsystems illustrating the control flow and the data flow. The feedback to the

control loop is modeled in the data flow when traversed in reverse from the bottom of the tree. The functional dependency tree is as shown in Figure 3.11. A functional dependency tree models the functions and their dependencies hidden to the user in a hierarchical manner.

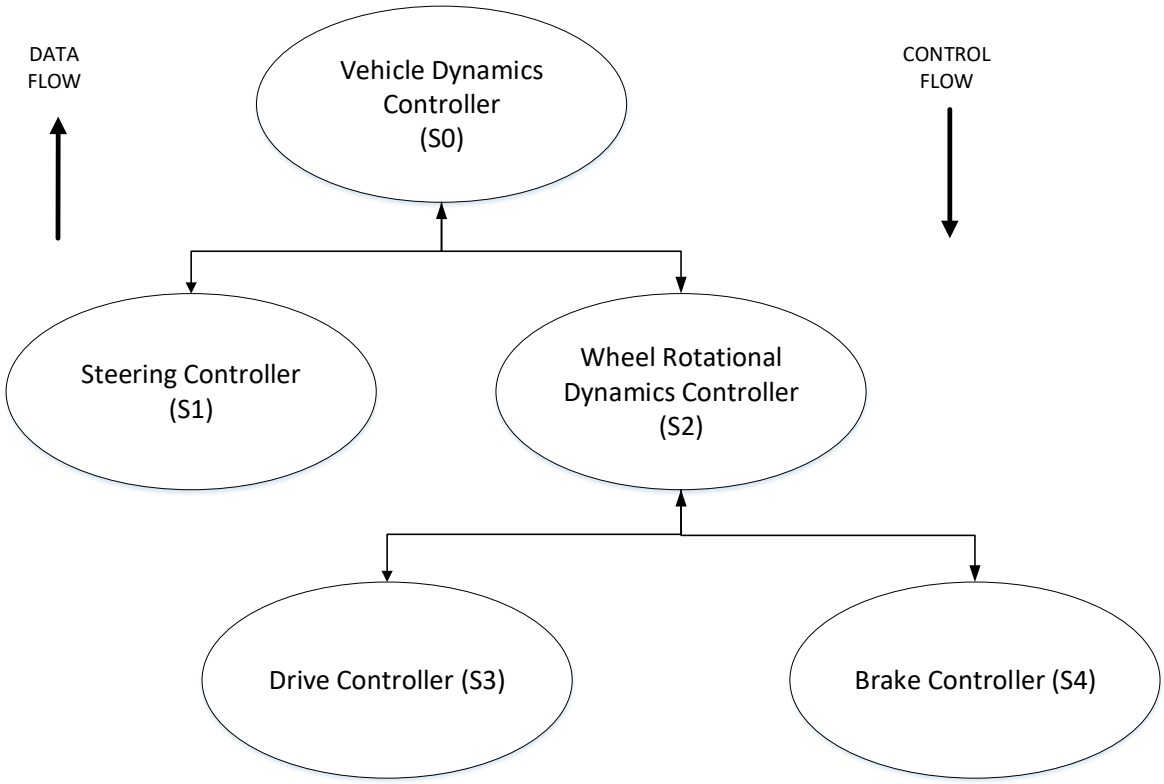

Figure 3.11: Step 1: Functional dependency tree of the actuation system

In step 2, we perform the hazard and risk analysis (HARA) prescribed in ISO 26262 for the system based on the abstraction level in the functional dependency tree. The top-level item, the vehicle dynamics controller, is chosen and the possible malfunctions experienced by it are enumerated for different operating situations. For every hazardous situation the severity and exposure are assigned. The risk is calculated in a quantitative manner as a product of the probability of harm and the severity of that harm. As an example, we present the HARA for the vehicle dynamics controller and the wheel rotational dynamics controller, and

choose one malfunction to analyze in Table 3.11 and Table 3.12. The expected severity level for most high speed accidents is S3 and is equal to a value of 4 on our scale where 1 is the lowest and 4 is the highest.

Table 3.11: Step 2: Hazard and Risk Analysis Results for Actuation System Part I

Item	Malfunction	Driving Situation	Risk Scenario	Hazard
Vehicle Dynamics Controller (S0)	f1: Incomplete/Missing Trajectory Information	Curved steep mountain roads at high speeds ($>$ 40 mph)	Loose gravel, incoming traffic	Collision possible
	f2: Error in controller steering data output	Freeway at high speeds ($>$ 40 mph)	Heavy obstacle present at $< 300m$	Collision possible
	f3: Error in controller output - wheel rotational dynamics data	Freeway at high speeds ($>$ 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f4: Erroneous/Missing controller feedback - Vehicle state	Freeway at high speeds ($>$ 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
Wheel Rotational Dynamics Controller (S2)	f5: Erroneous/Missing controller output	Urban roads at medium speeds ($<$ 40 mph)	Incoming traffic	Collision possible
	f6: Input to controller missing	Freeway at high speeds ($>$ 40 mph)	Heavy obstacle present at $< 300m$	Collision possible
	f7: Drive controller and Brake controller both receive values simultaneously	Freeway at high speeds ($>$ 40 mph)	Heavy obstacle present at $< 300m$	Collision possible
	f8: Error in controller output - Braking value sent to Drive controller	Freeway at high speeds ($>$ 40 mph)	Heavy obstacle present at $< 300m$	Collision possible due to sudden acceleration instead of braking
	f9: Missing feedback	Freeway at high speeds ($>$ 40 mph)	Heavy obstacle present at $< 300m$	Collision possible

Table 3.12: Step 2: Hazard and Risk Analysis Results for Actuation System Part II

Item	Malfunction	Driving Situation	Risk Scenario	Hazard
Braking Controller (S3)	f13: Incomplete/Missing input information	Curved steep mountain roads at high speeds (> 40 mph)	Loose gravel, incoming traffic	Collision possible
	f14: Error in controller braking torque data output	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 300m$	Collision possible
	f15: Braking torque output results NULL or UNKNOWN	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
Drive Controller (S4)	f16: Incomplete/Missing input information	Steep uphill mountain roads at high speeds (> 40 mph)	traffic following vehicle	Rear Collision possible
	f17: Error in controller engine torque output	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 300m$	Collision possible

For our current example, we analyze system S2 for the fault $f2$. assuming the probability of harm to be 0.7 and the severity to be a value of 3, we obtain $R = 2.1$ for the occurrence of fault $f6$ in subsystem S2. The safety engineer or designer sets the required residual risk R' = 0.8 for the system. The required residual risk is the maximum risk expected in a safe state of operation after recovering from a fault. In Step 3, we list the reactions available for the top-level system to mitigate the potential hazard due to the fault/malfunction. The existing reactions to the fault in analysis are listed in Table 3.13.

We determine the FRT and the FTTI values based on the vehicle body dynamics in step 4 and set the values for each subsystem in step 5 as shown in Figure 3.12. The risk mitigation factor for the fault in analysis is determined to be M3. The FDT is assumed to be 100 mS. The FTTI of the system is assumed to be 500 mS. Based on the FRT intervals, the system is assigned a risk mitigation factor of M3 as the FRT(fx, S0, S0) = 300 mS where

Table 3.13: Step 3: Reactions for fault $f6$

Reactions r_i	$T(r_i)$ mS	K_i
r_1: Change lane to left/right (S1)	200	0.4
r_2: Query for new inputs every 50mS (S2)	5	0.6
r_3: Halt - apply braking torque (S2)	200	0.2
r_4: Resend inputs (S0)	100	0.8
r_5: Estimate values based on wheel motion sensor feedback	350	0.7

Table 3.14: Step 5: Risk Mitigation Factor Windows

Risk Mitigation Factor Mx	Window W(Mx) in mS
M3	0 - 500
M2	501 - 1000
M1	1001 - 1500
M0	1501 - 2000

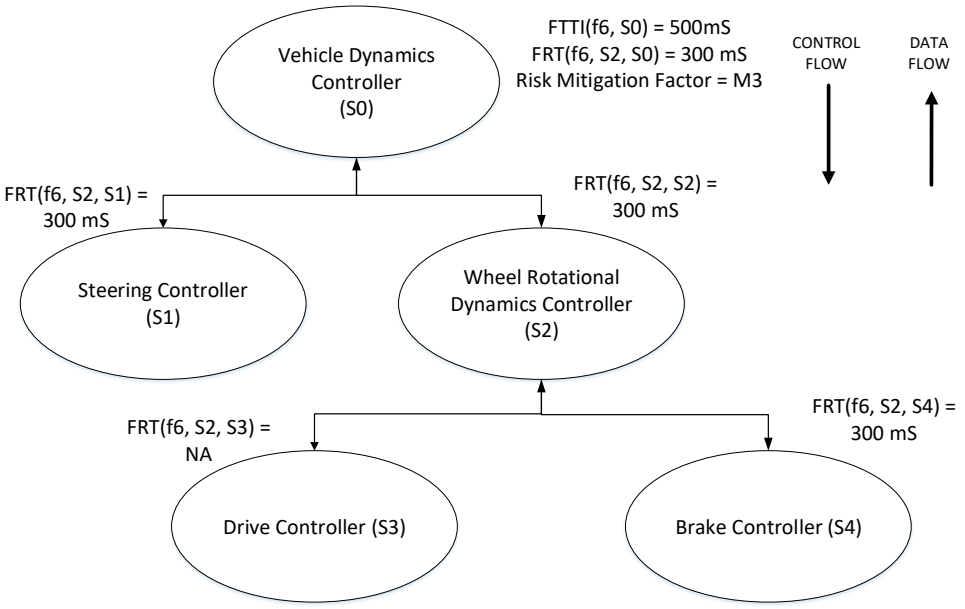

Figure 3.12: Steps 4 & 5: Determining FRT and Risk Mitigation Factor

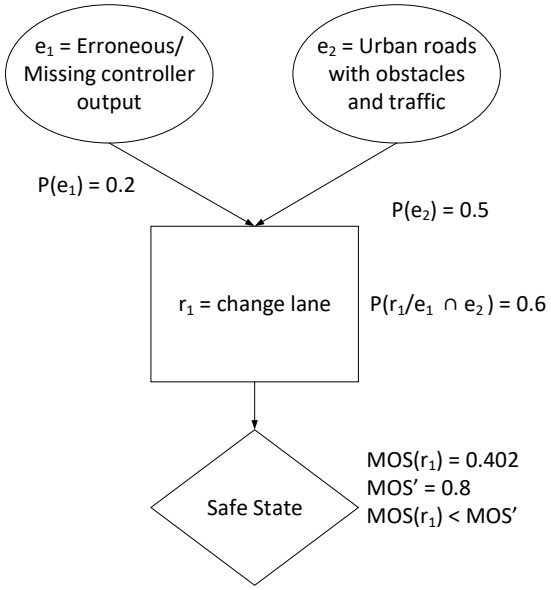

Figure 3.13: Step 6-6a: Influence diagrams for functional safety assessment

Table 3.15: Safety Goals For the System Derived from Influence Diagram Assessment

Safety Goal
FSG1: S1 MUST change lanes when S2 misses input values WITHIN T(r) < FRT(f5, S1, S1) = 300mS
FSG2: S2 MUST query for new inputs every 50mS WITHIN FRT(f6, S2, S2) = 300mS
FSG3: S3 MUST provide braking after change of lane when S2 misses input data and S0 is non responsive WITHIN FRT(f6, S2, S2) = 300 mS

fx represents the class of faults in analysis. For our current example the FRT values have been propagated top-down as shown in Figure 3.12 . In step 6, we determine the sufficient set of reactions r, that need to be incorporated into design by utilizing ID assessment and the measure of success $MOS(r)$ metric as shown in Figure 3.13. We observe for the given fault the reaction r_1 - change lane is insufficient. It does not reduce the risk satisfactorily in the given situation. The Hence, we reiterate steps 6-7 to obtain a satisfiable set of reactions.

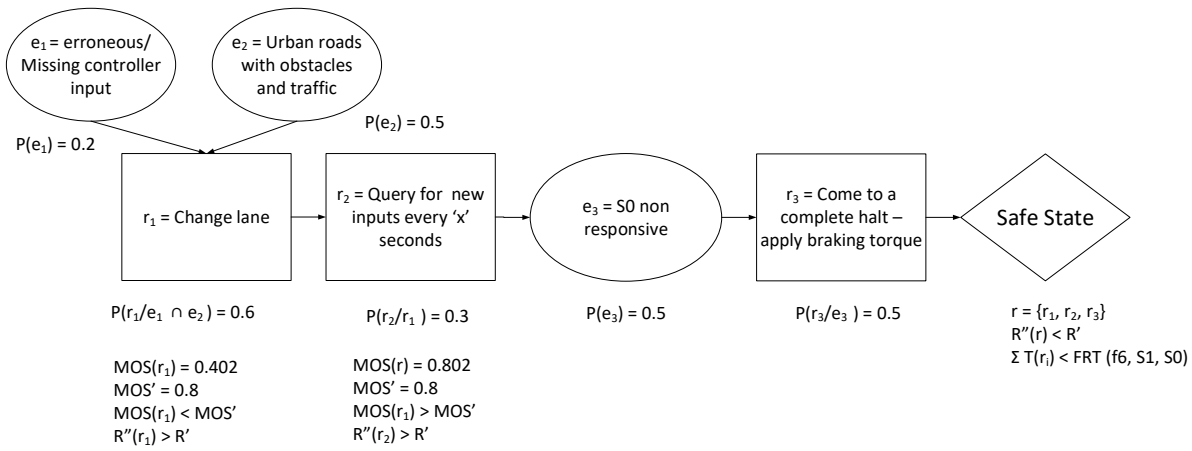

Figure 3.14: Step 7: Additional reactions added to obtain a set of reactions satisfying the satisfiability criterion

The designer sets the MOS' value = 0.8 as the threshold for the set of reactions r to ensure that the system satisfactorily reaches a safe state of operation. We obtain a set of reactions $r = \{r_1, r_2, r_3\}$ as shown in Figure 3.14 that satisfy our criterion. In step 9 as shown in Table 3.15.

We apply the same steps for the entire actuation system in analysis and obtain 24 safety goals shown in Table 3.17 and Table 3.18. Any redundant or superseded safety goal is removed. The AVIL for the system is found to be AVIL D for its development integrity level, which follows the process specified by the standard. These results are comparable to the results in the work presented by [26]. The authors have specified that the safety goals derived are for the actuation system only. The reactions existing and added are enforced as safety goals during the design process. The functional safety concept and the technical safety concept are derived in accordance to the prescribed methodology in ISO 26262. The advantage of FuSAV is clearly seen in analyzing fault f1. The fault f1 requires the extension of the described functionality of the Vehicle Dynamic Controller. When we perform the influence diagram assessment, we obtain the following safety goals which are comprehensive

for fault f1 shown in Table 3.16.

Table 3.16: Safety Goals For the System Derived from Influence Diagram Assessment for Fault f1

System	Safety Goal
S0	FSG1: S0 MUST Maintain trajectory copy in memory up to 600 seconds if S0 misses input values WITHIN T(r) < FRT(f1, S0, S0) = 300 mS
S0	FSG2: S0 MUST ASSURE memory protection for trajectory
S0	FSG3: S0 MUST request trajectory data every 300 seconds WITHIN FRT(f1, S0, S0) = 300 mS

Table 3.17: Complete Safety Goals For the System Level Actuation Faults Part I

System	Safety Goal
S0	FSG01: S0 MUST Maintain trajectory copy in memory up to 600 seconds if S0 misses input values WITHIN T(r) < FRT(f1, S0, S0) = 300 mS
S0	FSG02: S0 MUST ASSURE memory protection for trajectory
S0	FSG03: S0 MUST request trajectory data every 300 seconds WITHIN FRT(f1, S0, S0) = 300 mS
S0	FSG04: S0 MUST ASSURE availability of steering data within FRT(f2, S0, S0) = 100 mS
S0	FSG05: S0 MUST ASSURE availability of steering data within FRT(f3, S0, S0) = 300 mS
S0	FSG06: S0 MUST extrapolate from duplicate data present on redundant sensor(alternative wheel) within FRT(f4, S0, S0) = 300 mS
S0	FSG07: S0 MUST ASSURE feedback data within FRT(f4, S0, S0) = 300 mS
S1	FSG08: S1 MUST change lanes when S2 misses input values WITHIN T(r) < FRT(f5, S1, S1) = 300mS
S1	FSG09: S1 MUST ASSURE input availability & correctness WITHIN FRT(f10, S1, S1) = 100 mS
S1	FSG10: S1 MUST ASSURE steer angle output availability & correctness WITHIN FRT(f11, S1, S1) = 80 mS
S1	FSG11: S1 MUST ASSURE steer torque output availability & correctness WITHIN FRT(f12, S1, S1) = 80 mS
S1	FSG12: S1 MUST change lane when S3 is non responsive and obstacle detected < 200m WITHIN FRT(f15, S3, S1) = 300 mS

Table 3.18: Complete Safety Goals For the System Level Actuation Faults Part II

System	Safety Goal
S2	FSG13: S2 MUST recompute output values when missing values within FRT(f5, S1, S1) = 200 mS
S2	FSG14: S2 MUST ASSURE availability of input data within FRT(f6, S2, S2) = 300 mS
S2	FSG15: S2 MUST query for new inputs every 50mS WITHIN FRT(f6, S2, S2) = 300mS
S2	FSG16: S2 MUST prioritizes braking values when both controller values are simultaneously computed WITHIN FRT(f7, S1, S1) = 200 mS
S2	FSG17: S2 MUST ASSURE correctness of output transmission through system isolation WITHIN FRT(f8, S2, S2) = 300mS
S2	FSG18: S2 MUST ASSURE availability of sensor feedback through redundancy WITHIN FRT(f9, S2, S2) = 300mS
S3	FSG19: S3 MUST provide braking torque if S1 misses data for more than FRT(f10, S1, S1) = 100 mS within FRT(f10, S1, S3) = 300 mS
S3	FSG20: S3 MUST provide braking after change of lane when S2 misses input data and S0 is non responsive within FRT(f6, S2, S2) = 300 mS
S3	FSG21: S3 MUST ASSURE input data WITHIN FRT(f13, S3, S3) = 300 mS
S3	FSG22: S3 MUST ASSURE output data correctness when missing input data within FRT(f13, S3, S3)
S3	FSG23: S3 MUST provide redundant braking when NULL/UNKNOWN WITHIN FRT(f15, S3, S3) = 300 mS
S4	FSG24: S4 MUST ASSURE availability of input data WITHIN FRT(f16, S5, S4) = 600 mS
S4	FSG25: S4 MUST ASSURE drive torque output correctness WITHIN FRT(f17, S5, S4) = 300 mS

3.4 Case study: Decision & Control

We apply the FuSAV method on the decision and control system which acts as the brain of the system. The decision and control component of the architecture is responsible for path planning, determining the driving strategy, and the maneuver planning.

In Step 1 we construct the functional dependency tree to obtain a hierarchical relationship between the different subsystems illustrating the control flow and the data flow. The feedback to the control loop is modeled in the data flow when traversed in reverse from the bottom of the tree. The functional dependency tree is as shown in Figure 3.15. A functional dependency tree models the functions and their dependencies are abstracted from the designer.

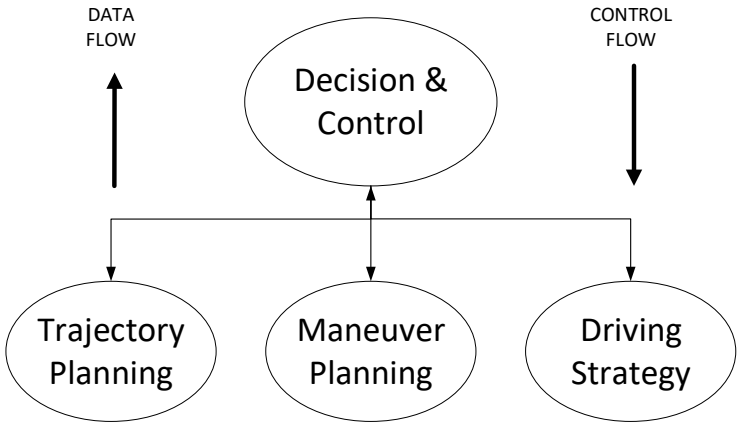

Figure 3.15: Step 1: Functional dependency tree

The top-level item, the decision & control module, is chosen and the possible malfunctions experienced by it are enumerated for different operating situations. For every hazardous situation the severity and exposure are assigned. The risk is calculated in a quantitative manner as a product of the probability of harm and the severity of that harm. As an

example of step 2, we present the HARA for the decision & control module, and choose one malfunction to analyze in Table 3.19. The interested reader can refer to the appendix for the complete table of top-level hazards and risks. The expected severity level for most high speed accidents is S3 and is equal to a value of 4 on our scale where 1 is the lowest and 4 is the highest.

Table 3.19: Step 2: Selective Hazard and Risk Analysis Results

Item	Malfunction	Driving Situation	Risk Scenario	Hazard
Decision & Control (S0)	f1: Missing/incomplete outputs	Curved steep mountain roads at high speeds ($>$ 40 mph)	Loose gravel, incoming traffic	Collision possible
	f2: Missing/incomplete sensor inputs	Freeway at high speeds ($>$ 40 mph)	Heavy obstacle present at $< 300m$	Collision possible
	f3: Erroneous sensor inputs	Freeway at high speeds ($>$ 40 mph)	Heavy obstacle present at $< 500m$	Collision possible

For our current example, we analyze system S0 for the fault $f1$. assuming the probability of harm to be 0.95 and the severity to be a value of 4, we obtain $R = 3.8$ for the occurrence of fault $f1$ in subsystem S0. The safety engineer or designer sets the required residual risk R' = 0.8 for the system. The required residual risk is the maximum risk expected in a safe state of operation after recovering from a fault. In Step 3, we list the reactions available for the top-level system to mitigate the potential hazard due to the fault/malfunction. The existing reactions to the fault in analysis are listed in Table 3.20. The events considered in analysis are listed in Table 3.21.

We determine the FRT = 100mS and the FTTI = 200mS in step 4 and set the values for each subsystem in step 5 as shown in Figure 3.16. The risk mitigation factor for the fault in analysis is determined in this step and assigned depending on the FRT range. We once again assume that the FDT is uniform for all faults and is known to be 100 mS. The window

Table 3.20: Step 3: Reactions for fault $f1$

Reactions r_i	$T(r_i)$ mS	K_i
r_1: Track safety shoulder	20	0.8
r_2: AV comes to a safe stop	100	0.2
r_3: Compute new outputs every 50mS	50	0.4
r_4: Reset the decision-making module	100	0.1
r_5: Estimate values based on previously known state	200	0.9

Table 3.21: Events affecting Reactions for fault $f1$

Events e_i
e_1: Missing outputs while on curved steep mountains at speeds >40 mph
e_2: System S0 non-responsive at speeds >40 mph

for the FRT's are divided in four equal intervals to obtain four windows as shown in Table 3.22. Based on the FRT intervals, the system is assigned a risk mitigation factor of M3 as the FRT(fx, S0, S0) = 100 mS. We propagate the FRT values top-down as shown in Figure 3.16 .

In step 6, we determine the sufficient set of reactions r, that need to be incorporated into design by utilizing ID assessment and the measure of success $MOS(r)$ metric as shown in Figure 3.17. The designer sets the MOS' value = 0.8 as the threshold for the set of reactions r to ensure that the system satisfactorily reaches a safe state of operation. As an example we consider that events e_1 and e_2 affect the system simultaneously with event e_2 being in effect after reaction r_1 completes i.e., a second reaction is needed to overcome the hazardous event. We obtain a set of reactions $r = \{r_1, r_2\}$ as shown in Figure 3.8 that satisfy our criterion. In step 9 as shown in Table 3.23, we specify the set of safety goals using the template shown in Figure 3.9. The Table 3.26 consists of all the safety goals obtained for the system when analyzed for the set of faults described in Table 3.24 and Table 3.25.

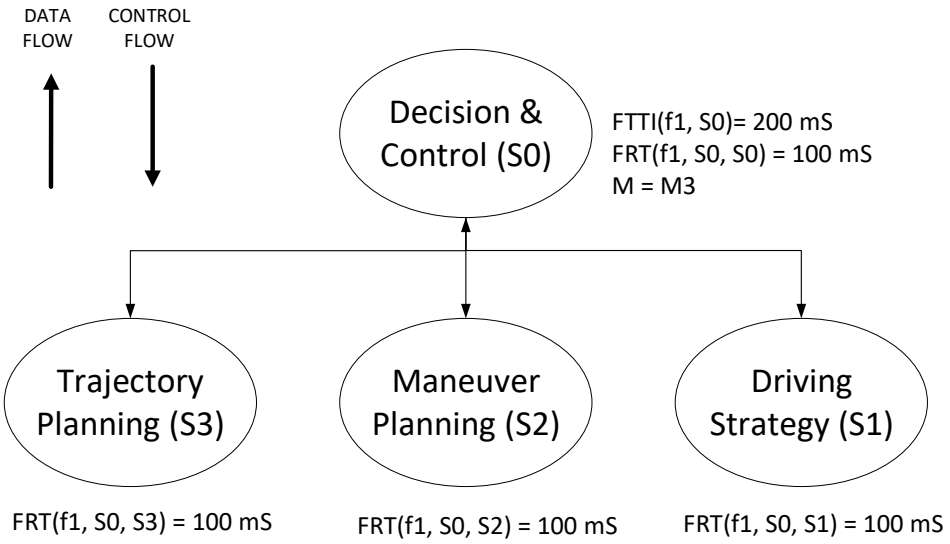

Figure 3.16: Steps 4 & 5: Determining FRT and Risk Mitigation Factor

Table 3.22: Step 5: Risk Mitigation Factor Windows

Risk Mitigation Factor Mx	Window W(Mx) in mS
M3	0 - 100
M2	101 - 200
M1	201 - 300
M0	301 - 500

Table 3.23: Safety Goals For the System Derived from Influence Diagram Assessment for Fault f1

Safety Goal
FSG1: S0 MUST track safety shoulder WITHIN T(r_1) < FRT(f1, S0, S0) = 100mS
FSG2: AV MUST come to a safe stop when S0 is non responsive WITHIN FRT(f1, S0, S5) = 100mS

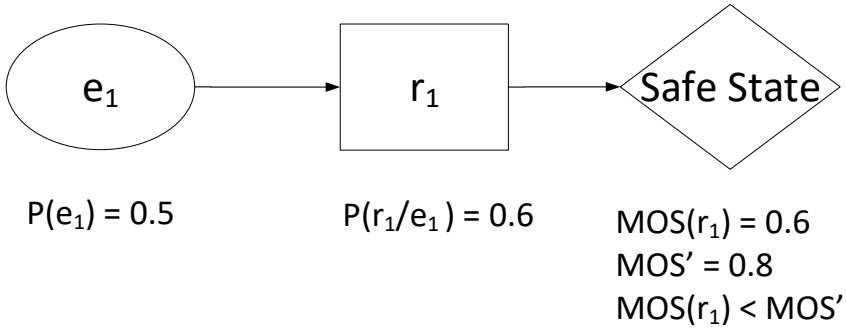

Figure 3.17: Step 6-6a: Influence diagrams for functional safety assessment

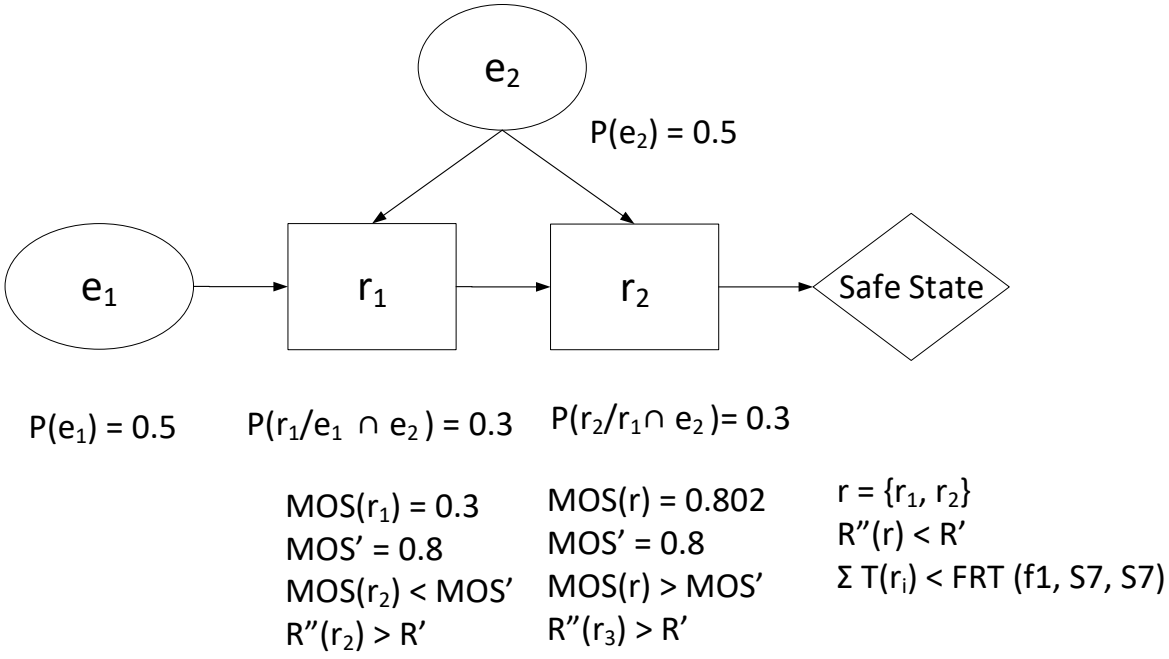

Figure 3.18: Step 7: Additional reactions added to obtain a set of reactions satisfying the satisfiability criterion

Table 3.24: Step 2: Hazard and Risk Analysis Results Part I

Item	Malfunction	Driving Situation	Risk Scenario	Hazard
Decision & Control (S0)	f1: Missing/incomplete outputs	Curved steep mountain roads at high speeds (> 40 mph)	Loose gravel, incoming traffic	Collision possible
	f2: Missing/incomplete sensor inputs	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 300m$	Collision possible
	f3: Erroneous sensor inputs	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f4:Error in trajectory computation	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f5: Error in driving strategy	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f6: Error in Maneuver Planning	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f7: Erroneous consensus	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f8: Delay in sensor inputs	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f9: Delay in computed outputs	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f10: Message corruption between actuation and decision-making	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible

Table 3.25: Step 2: Hazard and Risk Analysis Results Part II

Item	Malfunction	Driving Situation	Risk Scenario	Hazard
Decision & Control (S0)	f11: Incorrect interaction between driving strategy function and trajectory computation	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f12:Unintended interaction between actuation subsystems	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f13: Unintended interactions between sensing components	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f14: Incorrect interaction between decision-making elements	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f15: Delay in consensus	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f16: Loss in steering	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f17: Loss in braking	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible
	f18: Loss in engine function	Freeway at high speeds (> 40 mph)	Heavy obstacle present at $< 500m$	Collision possible

Table 3.26: Complete Safety Goals For the Decision & Control System

System Safety Goals
FSG1: S0 MUST ASSURE availability of computed outputs WITHIN T(r) < FRT(f1, S0, S0) = 100 mS
FSG2: S MUST come to a safe stop in case of non-responsive S0 WITHIN FRT(f1, S0, S5) = 200 mS
FSG3: S4 MUST track possible locations to stop WITHIN FRT(f1, S0, S4) = 100 mS
FSG4: S0 MUST Maintain trajectory copy in memory up to 'x' seconds if S0 misses input values WITHIN T(r) < FRT(f2, S0, S3) = 100 mS
FSG5: S0 MUST ASSURE memory protection for trajectory
FSG6: S8 System MUST ASSURE availability of redundant sensor data WITHIN FRT(f3, S0, S8) = 50 mS
FSG7: S0 MUST ASSURE availability of steering data within FRT(f4, S0, S0) = 100 mS
FSG8: S0 MUST verify driving strategy during route planning WITHIN FRT(f5, S0, S1) = 100 mS S0 MUST ASSURE correct maneuvering in any given situation WITHIN FRT(f6, S0, S2) = 100 mS
FSG9: S0 MUST ASSURE consensus of decisions WITHIN FRT(f7, S0, S0) = 30 mS
FSG10: S0 MUST ASSURE correct sensor feedback data within FRT(f8, S0, S0) = 100 mS
FSG11: S0 MUST ASSURE timeliness- correct output data within FRT(f9, S0, S0) = 100 mS
FSG12: S0 MUST ASSURE integrity of transmission between S0, S4 and S8
FSG13: S0 MUST transmit duplicated data along at least two redundant paths to S4 WITHIN FRT(f10, S0, S0) = 100 mS
FSG14: S0 MU4S T ASSURE isolation of non-interacting subsystems
FSG15: S0 MUST FSG15: ASSURE appropriate reaction for steering sub-system failure WITHIN FRT(f16, S5, S0) = 100 mS
FSG16: S0 MUST ASSURE FSG16: change of lane if obstacle present when braking sub-system fails WITHIN FRT(f17, S6, S0) = 100 mS
FSG17: S0 MUST steer away froFSG17: m any obstacles when engine subsystem fails WITHIN FRT(f18, S7, S0) = 300 mS

3.5 Discussion of the Case Study

From the results presented in Sections 3.2, 3.3, and 3.4, we notice that FuSAV method sufficiently can identify safety goals at different levels of abstraction. When compared with ISO 26262 and ISO 26262 with STPA in Section 3.2, the safety goals are comparable, with the added benefit that the set of reactions obtained in the influence diagram assessment reflect the capability of the vehicle to reach a safe state. The decision-safety paradigm and the quantification of risk allows designers to develop strategies to reduce the risk by leverage the system's ability to provide detrimental action to reach an acceptable risk level in a hazardous situation. These reactions need not necessarily involve high-level decision making by the vehicle. They could be in the form of passive safety mechanisms that provide fail-silent and fail-safe behavior for the different components and subsystems, while behaving in a fail-operational manner at the system level. The risk mitigation factor assess the AV for safety by considering both the probabilistic aspects of severity and exposure, as well as the deterministic nature of the vehicles real-time requirements. Since risk mitigation factor is highlighting the scope of the reactions and their timeliness, it is effective as a reproducible metric to assess the safety requirements. The safety goals obtained are also verifiable using the influence diagrams. We can certify that the system is safe if its behavior in simulated uncertain operating situations is as expected. The impact of the safety goal assessment method is clearly seen in the Section 3.4. The safety goals obtained cover all the listed types of malfunction. We do not claim that these list of safety goals obtained are complete and wholly sufficient as there are very few comparable studies for assessing the safety aspect of a fully autonomous car, instead we put forth an argument that the current methods can be enhanced by incorporated the risk mitigation factor and the influence diagram assessment presented in this work.

The FuSAV method does have its drawbacks. The most noted one being that constructing

influence diagrams and modeling all the uncertainties and the states of operation is computationally expensive. It also suffers from the drawbacks of depending on traditional HARA described in ISO 26262 – it is a brainstorming method and requires experience and insight to exhaustively list out all malfunctions. The reliance on probabilistic methods and quantifying risk also has its disadvantages. It may be that certain probabilities are unknown and for certain edge cases in driving scenarios the designer may not be able to assign a risk value. Despite all the mentioned drawbacks, the method allows the designer to model the uncertain driving environment of operating in a state of malfunction. The method also allows designers to construct reactions which can be implemented within the system safety.

Chapter 4

Conclusion & Future Work

In the current chapter, we discuss the potential future work in Section 4.2 and conclude the work presented in Section 4.1.

4.1 Conclusions

Safety goals are the top-level requirements that the AVs require to satisfy to assure safety in operation. The increasingly complex nature of these safety critical systems have compounded the difficulty of assessing for safety. The current edition of ISO 26262 is not sufficient by itself to assess for safety and requires a revision. With "Controllability" no longer being directly applicable as a usable metric for fully autonomous cars, we need to introduce newer metrics to assess and validate the development integrity.

The method described in the current work in Section 2.4 analyzes safety in the context of system-level decision-making by the AV and how it affects safety. Influence diagrams are capable of qualitatively and quantitatively assessing the decision-making capability of the AV in terms of safety with uncertainty. Influence diagrams allow us to model the driving environment, the failure rate of the components, and how certain reactions to a fault can ensure that the system maintains a safe state of operation.

From the influence diagram assessment, we are able to define a risk mitigation strategy that

defines how the vehicle should behave in a hazardous event.

4.2 Future Work

Future work includes defining a method to develop reactions in accordance to the decision-safety paradigm, and the determination of optimal risk mitigation strategies by minimizing the risk in the safe state. Another potential direction is in deriving verifiable models from the influence diagrams that can be used to assess the validation of safety goals at the different levels of abstraction.

Chapter 5

Summary

The safety goals and functional safety concept for vehicles are defined assuming that the driver is the redundant factor in a hazardous situation and is capable of mitigating the hazard through some action/maneuver. In a fully autonomous vehicle, the driver is a user of a service provided by the vehicle. The driver's responsibilities have been taken over by the vehicle and he is no longer part of the control loop. Hence the current version of the ISO 26262 is inadequate in holistically determining the safety requirements of the system – "Controllability" needs to be replaced and a method to evaluate the decision making capability of the vehicle in the absence of driver becomes necessary. Without such a defined metric, evaluating the functional safety and identifying the set of safety goals for an autonomous vehicle is difficult. Our contributions in the current work are as follows:

- We have defined a new fault reaction time based metric, Risk Mitigation Factor (M) to replace controllability used as an impact factor for functional safety assessment in autonomous vehicles.
- We present a method to evaluate the functional safety and derive the safety goals for autonomous vehicles.
- We present several case studies highlighting the effectiveness of the proposed approach.
- We utilize Influence Diagrams to model the AV's decision-making capability and its expected behavior in a hazardous situation. The optimal set of decisions by the AV

reduce the risk. Influence Diagrams model the system for safety in both a qualitative and quantitative manner.

- We quantitatively assess risk by assigning it a numeric value. The influence diagrams are used to minimize this numerical value by analyzing the driving situation and the vehicle's reaction to a malfunction.

Bibliography

[1] Asim Abdulkhaleq, Stefan Wagner, Daniel Lammering, Hagen Boehmert, and Pierre Blueher. Using stpa in compliance with iso 26262 for developing a safe architecture for fully automated vehicles. *arXiv preprint arXiv:1703.03657*, 2017.

[2] Algirdas Avizienis, J-C Laprie, Brian Randell, and Carl Landwehr. Basic concepts and taxonomy of dependable and secure computing. *IEEE transactions on dependable and secure computing*, 1(1):11–33, 2004.

[3] Sagar Behere and Martin Törngren. A functional architecture for autonomous driving. In *Proceedings of the First International Workshop on Automotive Software Architecture*, WASA '15, pages 3–10, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3444-0. doi: 10.1145/2752489.2752491. URL http://doi.acm.org/10.1145/2752489.2752491.

[4] Alessandro Birolini. *Reliability engineering*, volume 5. Springer, 2007.

[5] R Debouk, B Czerny, J dAmbrosio, and JJ Joyce. Safety strategy for autonomous systems. In *International Systems Safety Conference. System Safety Society*, volume 3, 2011.

[6] Clifton A Ericson and Clifton Ll. Fault tree analysis. In *System Safety Conference, Orlando, Florida*, pages 1–9, 1999.

[7] S. Geronimi, V. Abadie, and N. Becker. *Methodology to Assess and to Validate the Dependability of an Advanced Driver Assistance System (ADAS) Such as Automatic Emergency Braking System (AEBS)*, pages 125–131. Springer International Publishing,

Cham, 2016. ISBN 978-3-319-19818-7. doi: 10.1007/978-3-319-19818-7_13. URL http://dx.doi.org/10.1007/978-3-319-19818-7_13.

[8] Ronald A Howard and James E Matheson. Influence diagrams. *Decision Analysis*, 2 (3):127–143, 2005.

[9] M. Hrwick and K. H. Siedersberger. Strategy and architecture of a safety concept for fully automatic and autonomous driving assistance systems. In *2010 IEEE Intelligent Vehicles Symposium*, pages 955–960, June 2010. doi: 10.1109/IVS.2010.5548115.

[10] IEC. Functional safety of electrical/electronic/programmable electronic safety-related systems (e/e/pe, or e/e/pes). Standard, International Electrotechnical Commission, Geneva, CH, January 2010.

[11] ISO. Raod vehicles - functional safety - parts 1-10. Standard, International Organization for Standardization, Geneva, CH, January 2011.

[12] Rolf Johansson, Jonas Nilsson, Carl Bergenhem, Sagar Behere, Jörgen Tryggvesson, Stig Ursing, Andreas Söderberg, Martin Törngren, and Fredrik Warg. Functional safety and evolvable architectures for autonomy. In *Automated Driving*, pages 547–560. Springer, 2017.

[13] J. C. Laprie. Dependable computing and fault tolerance : Concepts and terminology. In *Fault-Tolerant Computing, 1995, Highlights from Twenty-Five Years., Twenty-Fifth International Symposium on*, pages 2–, Jun 1995. doi: 10.1109/FTCSH.1995.532603.

[14] Nancy Leveson. *Engineering a safer world: Systems thinking applied to safety*. MIT press, 2011.

[15] Helmut Martin, Kurt Tschabuschnig, Olof Bridal, and Daniel Watzenig. Functional

safety of automated driving systems: Does iso 26262 meet the challenges? In *Automated Driving*, pages 387–416. Springer, 2017.

[16] Robin McDermott, Raymond J Mikulak, and Michael Beauregard. *The basics of FMEA*. SteinerBooks, 1996.

[17] Mark L McKelvin Jr, Gabriel Eirea, Claudio Pinello, Sri Kanajan, and Alberto L Sangiovanni-Vincentelli. A formal approach to fault tree synthesis for the analysis of distributed fault tolerant systems. In *Proceedings of the 5th ACM international conference on Embedded software*, pages 237–246. ACM, 2005.

[18] Jamil K Naufal, João B Camargo, Lucio F Vismari, Jorge R de Almeida, Caroline Molina, Rodrigo Ignacio R González, Rafia Inam, and Elena Fersman. A^2cps: A vehicle-centric safety conceptual framework for autonomous transport systems. *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[19] NHTSA. Federal automated vehicles policy. https://one.nhtsa.gov/nhtsa/av/pdf/Federal_Automated_Vehicles_Policy.pdf, 2016.

[20] Josef Nilsson, Carl Bergenhem, Jan Jacobson, Rolf Johansson, and Jonny Vinter. Functional safety for cooperative systems. Technical report, SAE Technical Paper, 2013.

[21] Felix Redmill, Morris Chudleigh, and James Catmur. *System safety: HAZOP and software HAZOP*. Wiley Chichester, 1999.

[22] Andreas Reschka. Safety concept for autonomous vehicles. In *Autonomous Driving*, pages 473–496. Springer, 2016.

[23] Hans-Leo Ross. *System Engineering for Development of Requirements and Architecture*, pages 75–199. Springer International Publishing, Cham, 2016. ISBN 978-3-

319-33361-8. doi: 10.1007/978-3-319-33361-8_4. URL http://dx.doi.org/10.1007/978-3-319-33361-8_4.

[24] SAE. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. standards.sae.org/j3016_201609/, 01 2014.

[25] Ross D Shachter. Evaluating influence diagrams. *Operations research*, 34(6):871–882, 1986.

[26] Torben Stolte, Gerrit Bagschik, and Markus Maurer. Safety goals and functional safety requirements for actuation systems of automated vehicles. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, pages 2191–2198. IEEE, 2016.

[27] Torben Stolte, Ren S. Hosse, Uwe Becker, and Markus Maurer. On functional safety of vehicle actuation systems in the context of automated driving. *IFAC-PapersOnLine*, 49(11):576 – 581, 2016. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2016.08.084. URL http://www.sciencedirect.com/science/article/pii/S2405896316314343. 8th IFAC Symposium on Advances in Automotive Control AAC 2016.

[28] Sardar Muhammad Sulaman, Taimoor Abbas, Krzysztof Wnuk, and Martin Höst. Hazard analysis of collision avoidance system using stpa. In *Proceedings of the 11th International ISCRAM Conference*, 2014.

[29] Dajiang Suo, Sarra Yako, Mathew Boesch, and Kyle Post. Integrating stpa into iso 26262 process for requirement development. Technical report, SAE Technical Paper, 2017.

[30] Ömer Şahin Taş, Florian Kuhnt, J Marius Zöllner, and Christoph Stiller. Functional

system architectures towards fully automated driving. In *Intelligent Vehicles Symposium (IV), 2016 IEEE*, pages 304–309. IEEE, 2016.

[31] Fredrik Warg, Martin Gassilewski, Jörgen Tryggvesson, Viacheslav Izosimov, Anders Werneman, and Rolf Johansson. Defining autonomous functions using iterative hazard analysis and requirements refinement. In *International Conference on Computer Safety, Reliability, and Security*, pages 286–297. Springer, 2016.

[32] Wired. Gm will launch robocars without steering wheels next year, 2018. URL https://www.wired.com/story/gm-cruise-self-driving-car-launch-2019/.

[33] Junfeng Yang, Michael Ward, and Jahangir Akhtar. The development of safety cases for an autonomous vehicle: A comparative study on different methods. Technical report, SAE Technical Paper, 2017.

[34] B. Zheng, H. Liang, Q. Zhu, H. Yu, and C. W. Lin. Next generation automotive architecture modeling and exploration for autonomous driving. In *2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pages 53–58, July 2016. doi: 10.1109/ISVLSI.2016.126.