STPA FOR AUTONOMOUS VEHICLE SAFETY IN TRAFFIC SYSTEMS

Muhammad Rafay Ejaz
Michael Chikonde
DEPARTMENT OF ARCHITECHTURAL AND CIVIL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

Master's thesis 2022

STPA FOR AUTONOMOUS VEHICLE SAFETY IN TRAFFIC SYSTEMS

Muhammad Rafay Ejaz
Michael Chikonde

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Geology and Geotechnics
Chalmers University of Technology
Gothenburg, Sweden 2022

System Theoretic Process Analysis of Autonomous Vehicles in Traffic Systems

Supervisor: Kun Gao, Architecture and Civil Engineering
Examiner: Jiaming Wu, Architecture and Civil Engineering

Master's Thesis 2022
Department of Architecture and Civil Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2022

Abstract

It is speculated that vehicle companies such as Volvo, are actively developing autonomous vehicles and planning to introduce fully Autonomous Vehicle (AV) test fleets in the transport systems of Swedish cities. The gradual introduction of autonomous vehicles is meant to make the traffic system more efficient, reliable, and most importantly safer than a traditional human-driven vehicle (HDV) dominated system. However, the introduction of AVs will lead to an increase in the complexity of the traffic system. Especially the complex interactions between AVs with human drivers and other traffic participants such as pedestrians. These interactions are further complicated by operational domain constraints such as road infrastructure and natural weather conditions. These complex interactions create conditions that inevitably lead to unforeseen unsafe actions by a human or non-human participant, resulting in a hazardous scenario and an eventual accident event. Traditional hazard analysis methods as well as domain-specific ISO standards such as ISO 26262 mainly focus on AV component and functional failures, ignoring high-level safety hazard casual factors originating from system interactions and organizational flaws. New hazard analysis tools such as System Theoretical Process Analysis (STPA) provide systematic and sequential steps for identifying hazardous interactions within a system in the absence of extensive test trials and operational data. This thesis aimed to provide a pilot application of STPA for evaluating the systematic safety of AV in different traffic scenarios. The case study in this thesis is AV at an unsignalized intersection with an unprotected left turn.

The framework of the methodology in this thesis is formulated by using the STPA process. The first step defines the scope and objectives of the analysis. This involves defining the traffic system, enumerating its hazardous states, and identifying losses unacceptable to stakeholders. The second step establishes the traffic system control structure consisting of controllers, actuators, sensors, and controlled components. The third step identifies Unsafe Control Actions (UCA) that under the worst or extreme environmental conditions will lead to one or more hazards. Lastly UCA causal factors are identified through a brainstorming process. A hazardous Factor Network (HFN) is developed for a structured causal factor identification process. Further, principles of network analysis such as network traceability and betweenness values are utilized for causal factor (CF) evaluation.

The results show that STPA can be applied to identify hazard causal factors in a

traffic system with autonomous vehicles. The result also showed that hazard causal factors in a traffic system are highly connected and related to each other. The HFN provided a structured framework for casual factor traceability and evaluation. For example, identified CF5 (inadequate road infrastructure maintenance and management plan by road authority) had the highest betweenness value in the HFN. Meaning that CF5 is highly connected to other causal factors and UCAs, hence is very vital to the safety of AVs in the current traffic system. In conclusion, the thesis shows that in the absence of high statistical data, STPA can be applied as a framework for analyzing traffic system safety with the introduction of fully autonomous vehicles in the current system. This thesis finally recommends further research and development of the STPA process to include a "probabilistic link" between the Unsafe Control Actions (UCA) and established causal factors.

Acknowledgements

First and foremost, we would like to express our sincerest gratitude towards Kun Gao and Lei Chen. Both have played an instrumental part in the completion of this thesis where Kun Gao has provided us with vital support throughout the project not only at the academic level but on a personal level which is highly admiring. Dr. Chen has provided a helping hand in terms of understanding the basics of automation and safety and shared his inside knowledge from his special expertise in Intelligent Transportation Systems. This thesis would not have been possible without their consistent and timely support.

Muhammad Rafay Ejaz and Michael Chikonde, Gothenburg, June 2022

List of Acronyms and Abbreviations

ACC	Adaptive Cruise Control
ADAS	Advanced Drive Assistance Systems
NHTSA	National Highway Traffic Safety Administration
SAE	Society of Automotive Engineers
DDT	Dynamic Driving Task
ODD	Operational Design Domain
UCA	Unsafe Control Action
SC	Safety Constraints
CF	Causal Factors
AV	Autonomous Vehicle

Contents

List of Figures

List of Tables

1
Introduction

The introduction of autonomous and semi-autonomous vehicles into the current traffic system is becoming a reality with each passing day. Companies such as Volvo already have pilot projects such as "drive me", which aim to introduce at least 100 fully Autonomous Vehicles (AVs) in the city of Gothenburg (*Drive Me: Volvo Cars' approach to autonomous driving*, 2016). Soon, these initiatives will give birth to complex traffic systems consisting of tech instruments interacting with each other in several ways. The introduction of AVs in the current traffic system will inevitably give rise to new safety concerns beyond design flaws or component reliability. Technological advancements in the automotive industry have allowed vehicles to make autonomous decisions with little or zero human driver interference. A common notion in the road transport sector is that autonomous vehicles are safer than conventional vehicles. This is backed by statistical data showing that almost 95 percent of road accidents result from human error (Singh, 2015). However, accidents also occur due to unforeseeable interactions between different participants in the traffic system (Tingvall & Haworth, 1999). This is potentially truer for a traffic system with both conventional and fully autonomous vehicles as to be the most likely case in the near future. It is, therefore, vital that the safety of such a traffic system is explored further.

Various analysis methods have been developed for safety analysis since the 1990s. Currently, the automotive industry is fixated on the functional safety of autonomous vehicles through adherence to standards such as ISO 26262. This is usually coupled with common sequential event chain safety analysis tools such as the fault tree analysis and hazard analysis models such as the Swiss cheese model which views accidents as events waiting to happen (Hughes, Newstead, Anund, Shu, & Falkmer, 2015). These methods tend to omit hazard causation factors originating from different system levels due to complex system interactions, and external influences such as infrastructure and weather. When analyzing hazard causation using traditional methods, vast tests or historical data are usually required. Therefore, there is a need for new hazard analysis models that can be utilized for safety analysis in the absence of big data. A new framework such as System Theoretical Process Analysis STPA attempts to solve this problem. STPA is based on systems theory and views safety as a system-level problem rather than focusing on individual components (Salmon, McClure, & Stanton, 2012). Unlike the previously mentioned safety models, STPA is applicable when limited information is available such as during early stages of system development.

This thesis is systems-thinking-inspired and attempts to apply the STPA safety analysis framework in identifying potential safety hazard aspects of introducing AVs in the current traffic system usually overlooked by more traditional safety analysis methods . The objective is to develop a framework for the application of STPA in traffic systems. This thesis also attempts to further evaluate identified hazard causal factors from the STPA process by use of network theory principles.

2

State of the art, research gap and objectives

2.1 Vehicle Automation

In current road cars, the level of automation is rising. Advanced Driver Assistance Systems (ADAS), Adaptive Cruise Control (ACC) and other advanced features are becoming more common today. As a result, the prospect of fully driverless vehicles in the future is becoming more conceivable (NHTS, 2013). The extent of cars automation can be measured using a variety of indicators.

Driving maneuvers that were previously done by humans are being transferred to new driving systems as vehicles become more automated. As a result, classifications describing the various degrees of driving automation are used to distinguish vehicle automation levels. The US National Highway Traffic Safety Administration (NHTSA), the German Federal Highway Research Institute and the Society of Automotive Engineers have each developed their own taxonomies for the classification of vehicle automation. They are very similar to each other; however, the Society of Automotive Engineers (SAE) definition is widely popular.

The On-Road AV Standards Committee of SAE International has produced the SAE standard J3016, which relates to the definition and characterization of AV terminologies. The SAE standard utilises the On-Road Motor Vehicle Automated Driving Systems, which highlights six automation levels (SAE, 2014). The SAE standard J3016 will refer to levels of automation in this thesis. The following are the six stages of automation described by the standard:

- **Level 0: No Automation**
  At this level, the human driver is driving and monitoring the surrounding areas.
- **Level 1: Driver Assisted Automation**
  Driver assistance system oversees either horizontal or longitudinal vehicle control at this level. However, the human driver keeps an eye on the road and is ready to act if necessary.
- **Level 2: Partial Automation**
  Both lateral and longitudinal control is automated at this level, although the human driver oversees the driving environment and is ready to act.
- **Level 3: Conditional Automation**

This level automates lateral and longitudinal control and the immediate driving environment. However, the driver is ready to take over at any time.

- **Level 4: Highly Automated**
  At this stage, the vehicle can operate autonomously under particular roads and environmental circumstances without human intervention.
- **Level 5: Fully Automated**
  At this level, the vehicle is in complete control of all aspects of driving and monitoring, regardless of the road or environmental conditions.

Levels of automation are based on the functions of the driving automation system and the human driver and operational design domain as follows:

1. **The Dynamic Driving Task, (DDT)**
2. **The backup method, Dynamic Driving Task (DDT)**
3. **The Operational Design Domain (ODD)**

2.1.1 Dynamic Driving Task

The Dynamic Driving Tasks includes maneuvering and control levels but excludes strategic level according to Michon (1985). Activities at the strategic level involve organizing and carrying out journeys from one point to another. Information processing is only sometimes required at intervals of a few minutes to many hours. The choices made here act as inputs for the level above. Tasks involving interactions with the environment and other road users go under the maneuvering or tactical level.

2.1.2 Operational Design Domain

The particular circumstances under which a driving automation system is intended to operate are included in the Operational Design Domain (ODD). These restrictions on location, travel routes, the environment, traffic, and speed are a few examples. For instance, an automatic driving system might be built to only be used on highways, in the 20-60 km/h range, with heavy traffic, only during the day, and without heavy rain.

2.2 Hazard Analysis For Autonomous Driving

Traffic systems are increasingly becoming more complex with the introduction of autonomous driving. Traffic systems now not only comprise of human controllers but also computer algorithms. In a traffic system, different stakeholders and traffic participants need to interact in with each in an operation domain limited by infrastructure and natural weather conditions. For such a systems to operate safely, extensive hazard analsysis is vital. Accident cannot be solely viewed as originating from system component failure but also from complex interactions within a system.

There are various hazard analysis models and techniques for autonomous driving available. However, a comprehensive traffic safety analysis should include both

functional safety and the Safety Of The Intended Function (Zhang et al., 2010).Figure 2.2 illustrates the life cycle of automotive safety. Functional safety is related to potential hazards caused by failure of the system components to operate as per intended design. At the same time, Safety Of The Intended Function (SOTIF) is related to potential system hazards that arise minus any system component failure. The International Organization for Standardization addresses these two safety concepts through ISO 26262 and ISO 21448, respectively.

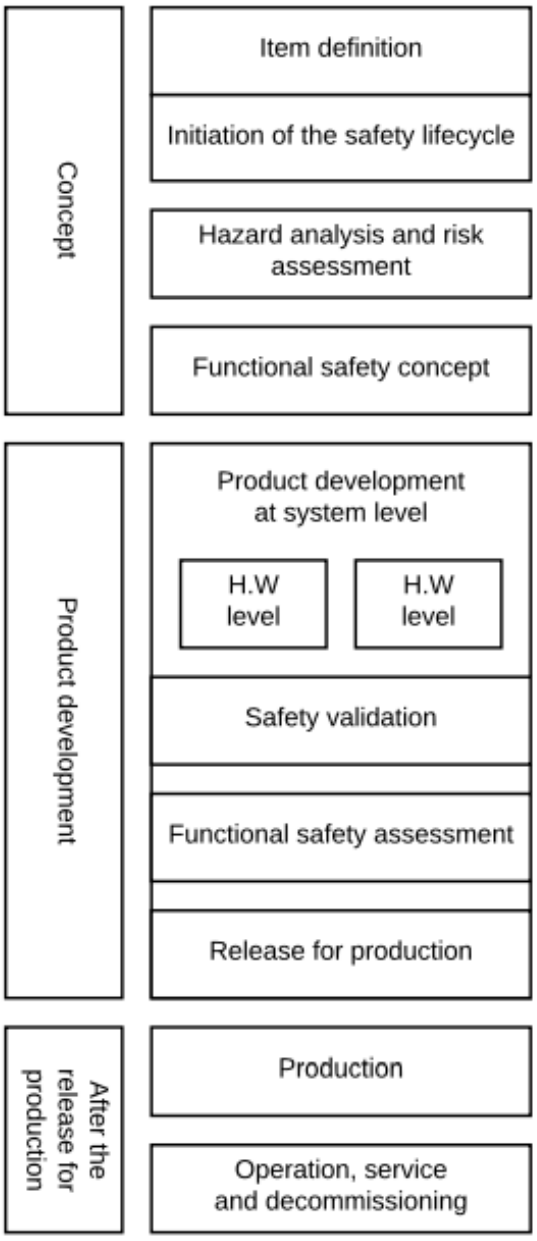

Figure 2.1: Life cycle of Automotive Safety (Zhang et al., 2010).

2.2.1 Autonomous functional safety

The importance of functional safety in the automotive industry cannot be overstated. While automobiles used to be entirely mechanical, they increasingly rely on Electrical and Electronic (E/E) systems (Zhang, Li, & Qin, 2010). As a result, embedded E/E systems and software incorporate more safety-critical functionalities. The ACC, anti-lock braking systems, collision warning systems, and airbag systems are some of the major features and functions of modern cars. The embedded systems that handle these functions are likewise considered safety-critical, and special precautions must be taken when working with them.

Standards ensure that a product is suitable for its intended usage. To assure safety, reliability, and quality, they establish requirements, specifications, recommendations, and characteristics. ISO (International Organization for Standardization) is a standard-setting organization that produced ISO 26262 in addition to the functional safety report (ISO, 2011). This standard specifies the requirements and procedures for dealing correctly with operational safety in automobile (E/E) systems. Adhering to this standard is the best available practice in the automotive sector. Before the advent of ISO 26262, there were multiple standards covering different areas of functional safety for vehicle E/E systems. ISO 26262 is based on previous automotive standards and research.

The standard is divided into ten sections as shown in Figure 2.2 covering safety management, product lifecycle management, analytical techniques, and safety principles. The product lifetime is separated into three phases: concept, product development, and post-release production. The concept definition includes item definition, risk assessment, hazard analysis, and functional safety. Similarly, the product development at the hardware level, functional safety review, safety validation, and final release for production are all part of the system-level product development. The post-production phase includes many activities, such as manufacturing, operation, servicing, and decommissioning. The important thing here is that STPA can be applied to all three stages of the life cycle development of an automobile.

Several authors have explored AV problems directly connected to ISO 26262 and its flaws. One major drawback is that the standard's scope limits dangers produced by electrical, electronic, and programmed electronic device faults. Others have questioned the structural suitability of the ISO 26262 standard for applicability to AVs and its scope constraints. Koopman and Wagner (2016) argues that current safety standards are insufficient for AV development, citing novel technologies and the lack of humans in the loop as essential factors to the complexity. As a result, Williams (2015) propose new modeling tools that holistically represent the interdependencies in autonomous systems.

2.2.2 Safety Of The Intended Function (SOTIF)

SOTIF (ISO 21448) was developed in 2021 with the aim of addressing the issues of functional insufficiency that ISO 26262 fails to address.Unlike previous functional

safety standards, which primarily focus on mitigating risks,SOTIF provides guidance to the design teams at a higher level, making sure that the design and validation is done taking into account all the possible risks that can arise in the real and complex environment.This makes car manufacturers to collect vast amounts of data and run huge number of simulations so that unknown scenarios can be identified and how car will perform under such scenario can be accurately measured.

SOTIF applicability is highly dependent on the nature of operional design domain. Major steps involved in the SOTIF analysis are defining the acceptance criteria for the unknown hazardous risks and secondly the validation targets that car manufacturers would try to achieve to meet those acceptance criteria through testing and data analysis by using different tools and strategies.little research is done in defining the acceptance criteria for the car manufacturers under various degrees of design domain.

2.2.3 Traditional Hazard Analysis Techniques

Some of the most commonly used methods in hazard analysis include Failure Mode Effect Analysis FMEA and Fault Tree Analysis FTA (Hughes et al. (2015)

2.2.3.1 Failure Modes and effect analysis

Failure Modes and Effects Analysis (FMEA) is a safety hazard analysis method used for identifying potential component (or subsystem) failures and assess the consequences of those failures (Pereira, Lee, & Howard, 2006) (J. Vincoli, 2006). The FMEA technique was created in the 1950s by reliability engineers working on military equipment.

Since then, defense, automotive, and other industries have embraced and adapted it. There are two categories of FMEA. The first is the inductive approach, where failure modes are known and their effects are identified. This strategy is mainly used for hardware faults and failures. The other is the deductive approach which deduces which modes can lead to failure. The following steps identify all the components involved within the system and their respective failure modes. After that, the effects of each loss are recorded, affecting other parts or sometimes the overall design.

The FMEA technique helps identify and characterize single-point failures in a system. The method's simplicity is advantageous since it can be used for a wide range of procedures at all stages of the design process. However, it is not well equipped for the evaluation of scenarios that can occur in complex, interconnected systems (Leveson, 1995).

2.2.3.2 Fault tree analysis

Fundamentally, Fault Tree Analysis (FTA) uses boolean logic focusing on the cause of events, resulting in a specific undesirable event leading to a hazard. If the analysis is qualitative, the result gives a list of possible combinations of circumstances,

including different events, environmental conditions, failures, that will result in danger. In contrast, if the data is quantitative, FTA will calculate the probability of the threat occurring within a particular period. FTA is widely used in risk analysis across many industries focusing on component reliability without considering the indirect factors that can affect the component being analyzed.

2.2.4 System theory perspective

Due to limitations of traditional hazard analysis techniques in coping with more complex systems,accident causation models based on a systems theory approach began to emerge in the 1930s and 1940s (Larsson et al. 2010).System theory focuses on systems as a whole and not on separate parts. This is the foundation for systems engineering. In systems engineering a system is viewed as an integrated whole even if it is composed of components each of a different nature. Currently the most popular accident causation models based on system theory are the Swiss Cheese model, Rasmussen's framework of risk management, and STAMP model (Salmon et al. 2012).

For the Rasmussen's risk management framework safety hazards are considered as resulting from organisational interactions between different hierarchical layers. In the Rasmussen's risk management framework, the upper levels elements impose decisions on the lower levels elements and the lower elements levels provide feedback to the higher levels elements (Salmon et al. 2012). Figure 2.3 illustrates Rasmussen's risk management framework

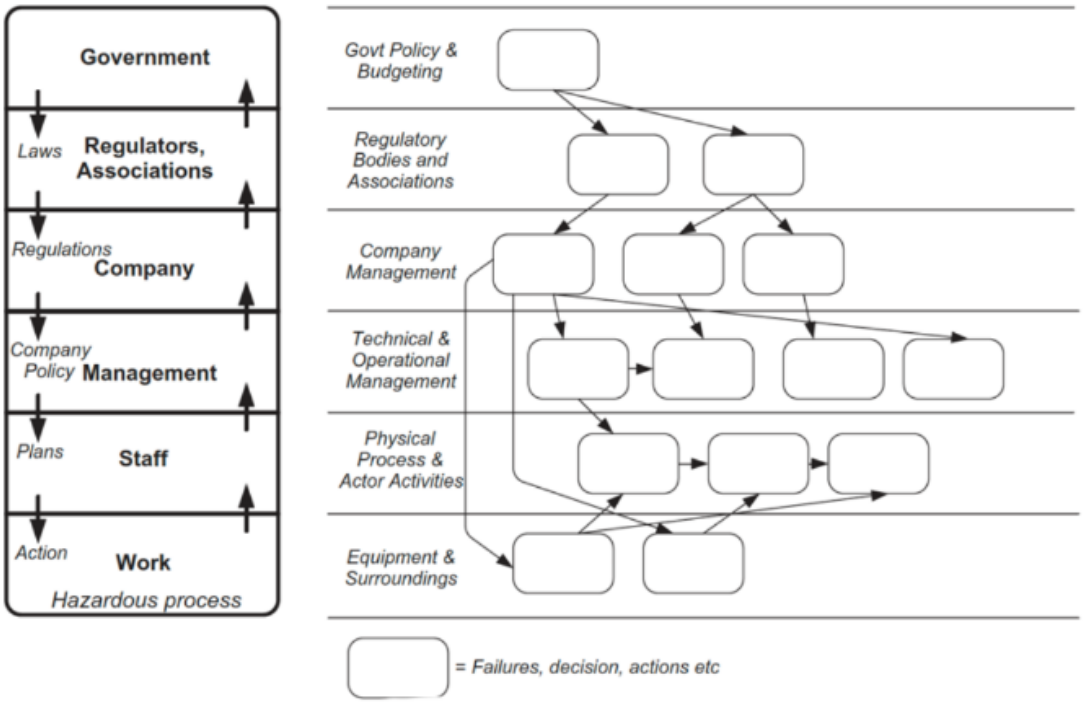

Figure 2.2: Rasmussen's risk framework (Salmon et al. 2012)

The Swiss Cheese model views organisations as comprising of multiple slices stacked side by side like cheese. Risk in the Swiss cheese model is mitigated by the adding layers of defenses behind each other. The main aim is risk prevention before any accident. This is illustrated in figure 2.4 below.

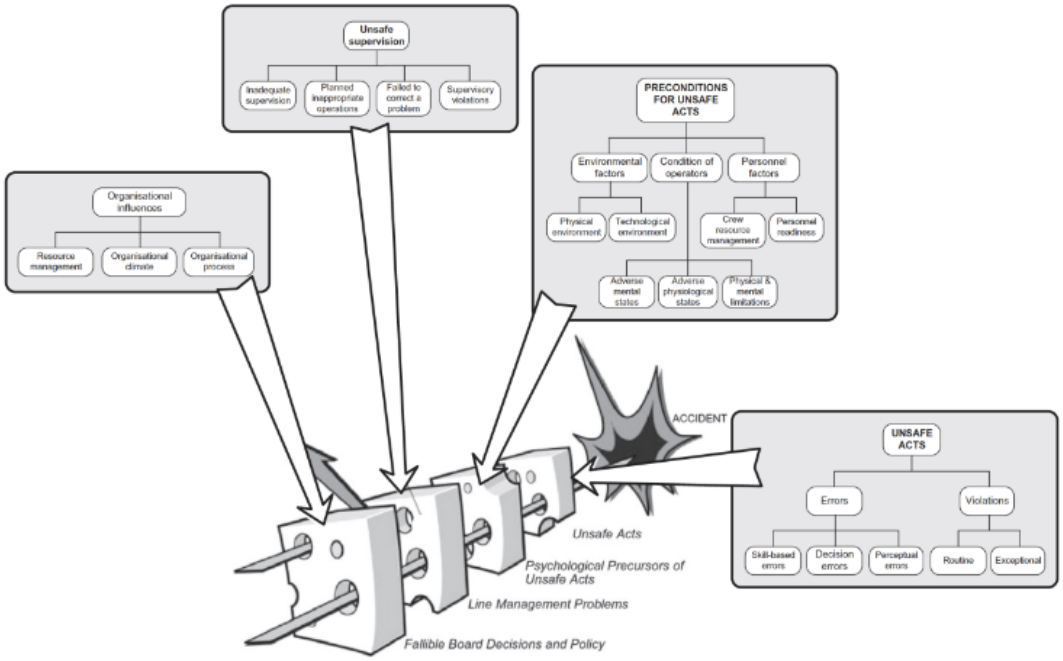

Figure 2.3: Rasmussen's risk framework (Salmon et al. 2012)

The STAMP model views systems as a hierarchical control processes. In the STAMP model higher-level components establish constraints on the lower-level components. The lower-level components, in turn, give feedback to the higher-level components. This is illustrated in figure 2.5 below. This STAMP model involves describing the system's control structure and identifying the failures leading to the accident.
Whilst the Swiss Cheese model and Rasmussen's risk management framework are model chain models where accidents are said to occur because of a sequence of failures of components, STAMP takes into consideration the relationships, or constraints that exist between different components of the system, and identify the constraints failures.With STAMP, the various components, their hierarchies, and constraints are all viewed together as a control system. This makes STAMP more suitable for more complex systems such as a Traffic System with Autonomous Vehicles.

2.2.4.1 System Theoretical Process Analysis (STPA)

System Theoretical Process Analysis (STPA) is based on STAMP and was designed specifically for analysing real-world systems. STPA focuses on analyzing the complex interactions and dynamic behavior of systems. In this was, STPA is advantageous over the traditional hazard analysis methods. It uses a system engineering approach

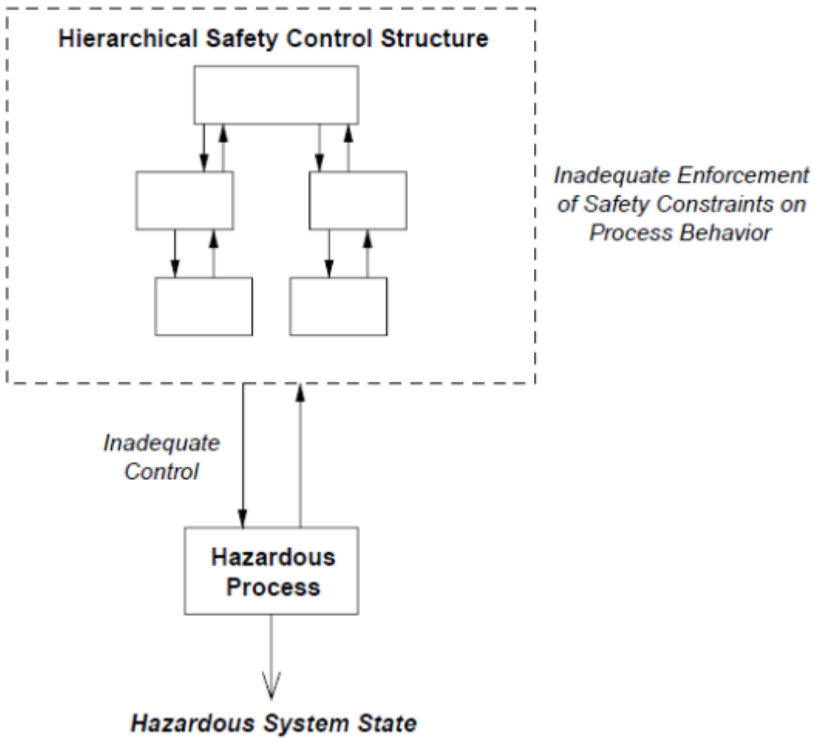

Figure 2.4: Hierarchical Control structure (Thomas et al., 2012) .

through the analysis of interactions within such a system.The STPA technique is a top-down strategy. According to Thomas (2012) STPA treats accidents as control challenges to ensure Safety Constraints (SCs) by managing the overall behavior of the system. Accident causality does not lend itself to a straightforward graphic portrayal, says Leveson (2007) as previous approaches frequently do. Figure 2.4 depicts a safety control framework with insufficient safety controls.

In the STPA process, the system's accidents and hazards are defined first. Nancy Leveson explain accidents and hazards in the following way:

- An accident is an unforeseen and undesired event which results in a loss
- A hazard is a state of the system or a combination of different conditions that will result in an accident when combined with the worst-case set of environmental variables.

The inverse of the identified hazards become the systems high-level constraints. Safety in the STPA approach is viewed as a control problem rather than a component failure problem because STPA is based on systems engineering. In the model, a control system can be described as components receiving feedback and acting on processes via actuators or sensors. Similarly, the elements that perform the control actions are known as controllers. Controlled processes are the systems they are in charge of, and the information that the controllers utilize to make choices is referred to as the process models. Control actions are the signals sent by the controllers.The final step in the STPA process is identifying Unsafe Control Actions and their respective casual factors.The results are then translated into system safety

requirements and constraints.

Attempts have been made to use STPA in safety analysis for various engineering and non-engineering systems. STPA has already been used to solve real-world problems. Laracy and Leveson (2007) used STAMP model in conjunction with STPA to assess the risks arising in critical infrastructure systems like air transportation networks. The LEX Comair 5191 was examined by (Nelson, 2008), where a plane crash investigation was conducted using the STPA model, and a causal factor breakdown was performed to identify root causes. STPA was found to be capable of capturing the complex nature of the accident, which was analytically consistent with reality. This application of STPA demanded that the technique include dangers and causes resulting from complex system interactions and lead the analytical process. By thoroughly addressing the entire system, including software, hardware, and operators' practices, and concentrating on the factors that have the most impact on safety, STPA was enacted. In other situations, one of which is the investigation of a railway accident in China, STAMP and STPA have proved successful (Ouyang et al., 2010). Other notable examples of STPA includes a safety analysis of the US Military Missile Defense System (Pereira et al., 2006). In the automotive industry, STPA have be used for hazard analysis of autonomous vehicle systems such as Lane keeping system,cruise control system, collision avoidance systems etc.

2.3 Research Gap and Objectives

Although there is a lot of research on the application of STPA for safety analysis of autonomous vehicles, however in literature, STPA has been has been applied only for safety hazard analysis of particular automation systems e.g lane keeping systems and cruise control systems. Thereby limiting the STPA analysis to describing the functioning of the vehicle controllers.There is a need to use STPA to explain the relationships of a System comprised of different sub-systems such a traffic system consisting a mixture of fully autonomous vehicle and traditional conventional vehicles. No attempts from literature were found where STPA was applied for safety analysis involving the entire traffic system.

In order to address these aspects, this thesis attempts to apply in safety hazard analysing of introducing a fully automated vehicle into the current traffic system. The following are the thesis objectives:

- To provide a pilot application of STPA for evaluating the safety of fully autonomous vehicles in different traffic scenarios within the current traffic system.
- To evaluation hazard causation factors and recommend improvement to application of STPA in safety analysis of traffic systems.

3
Methodology

Following the objectives set in the previous chapter, this chapter formulates the STPA approach. The process involves the theoretical description of the system and extensive hazard analysis using the STPA model. It also provides a structured casual factor identification process and further causal factor evaluation by introducing some principles of complex networks.

3.1 The STPA Approach

A-STPA software and excel was applied for this process.The STPA approach is based on the STAMP model and will be used as the basis of the framework for formulating the methodology. The STAMP model is convenient for understanding a System's components and sub-components of similar or different nature and their interactions with each other in several ways. In the context of this thesis, an autonomous vehicle in a trafic system will have to interact with other traffic participants and several external factors inclusing weather conditions and human behaviour. STPA provides systematic and sequential steps for conducting hazard analysis of such a System. The STPA approach essentially consists of 4 steps shown in Figure 3.1 below.

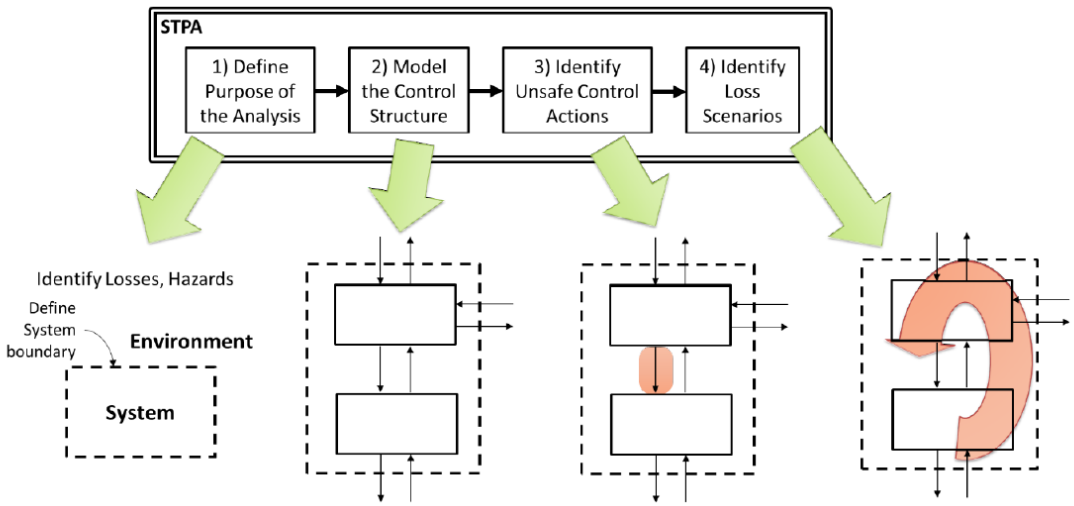

Figure 3.1: Overview of the basic STPA Method (France,2017)

1. Defining the purpose of the analysis: This is the preliminary step in the STPA framework and involves defining the scope and objectives of the analysis. This en-

tails defining the system under consideration and enumerating its hazardous states and losses.Hence, along with defining the System, the losses, hazards and constraints are identified.

2. Modelling the Control Structure: This is the second step in the STPA framework and it involves developing the control structure. In the STPA approach, a System is viewed as a control problem, with the higher-level controllers enforcing constraints on lower level controllers and components to achieve a particular goal.This step involves constructing the entire System, its controllers, components, sub-components, and how they interact with each other. The control structure typically includes controller which are the decision making components, actuators, sensors, the controlled process which are conceptual constructs of the beliefs that the controller has about the functioning of the process it is controlling. These components in the control structure interact through control action and feedback processes.

3. Identifying the Unsafe Control Actions: This step involves a procedure of identifying control actions and the various possible scenarios where under the worst or extreme environmental conditions, will lead to one or more Hazard. Control actions that lead to these scenarios are the unsafe control actions.The procedure for identifying unsafe control actions considers the scenarios such as not providing a Control Action, providing a control action too early/late late, stopping a control action too early/prolonged application or providing a wrong control action.

4. Identifying Loss Scenarios: This is the Final step in the STPA process and involves identifying the various possible causes leading to each of the identified Unsafe Control Actions in a systematic way. These results formulation the safety requirements for each specific causes/loss scenario.

3.1.1 Defining the purpose of the analysis

3.1.1.1 The system

In STPA a system can be defined as a group of interacting elements that act according to a set of rules to form a unified whole. The first step in the STPA analysis is defining the system under consideration. In this thesis a system traffic agents with autonomous driving vehicles as the focal point is taken under consideration. The elements that constitute the system are therefore those that directly or indirectly affect the safety performance of an autonomous vehicle in the current traffic system. Literature review and brainstorming is used to identify the constituents of the traffic system under consideration.

3.1.1.2 Losses/accidents

The STPA approach takes a top-down approach to meet the systems-theory methodologies. This is by firstly identifying unacceptable system losses. These are also known as accidents.

The losses are identified by taking into consideration situations that stakeholders deem unacceptable. Stakeholders in the this case includes all road users, regulators and vehicle manufacturers.

3.1.1.3 Hazards

The system hazards must be identified to describe the system-level requirements. These are the states of the System, which under extreme (or the worst) environmental conditions, will lead to one or more of the identified Losses in the first step of the STPA analysis .This concept of a hazard emphasizes the relevance of controllability and environmental risks since safety is achieved through control. The hazards must be defined so that they can be controlled by the designer or the operator, through design or operation. Nothing can be done to prevent a hazard if it is uncontrollable. Therefore, the system boundaries must be examined, or the hazard recast to fit within the controllable space in this situation.

3.1.2 System Control Structure

The control structure is a system's preliminary process model. In the STPA approach, the control structure is hierarchical in it's nature where high level elements enforce constraints on the behavior of lower level elements as shown in figure 3.2.

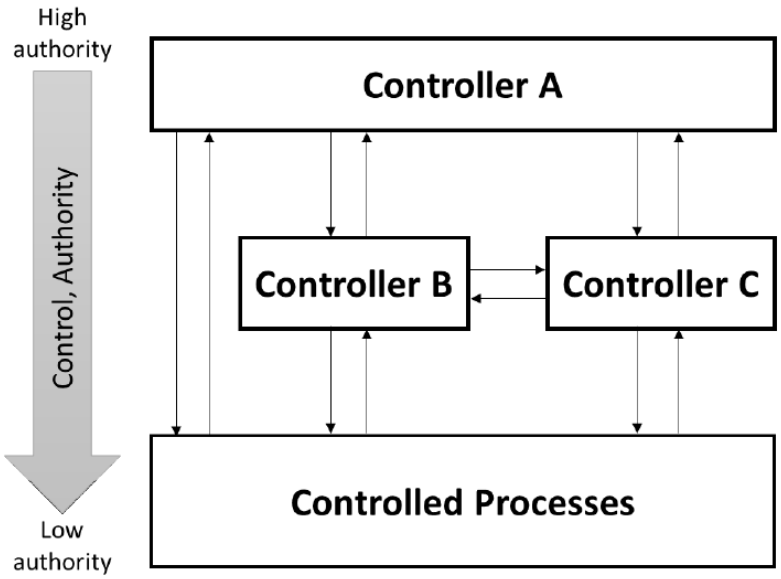

Figure 3.2: Generic hierarchical control structure (France,2017)

A hierarchical control structure generally contains Controllers, Control Actions, Feedback and the Controlled Process. Other components as in the case of a traffic system can be actuators and sensors. In an ideal control structure,the control algorithm is embedded inside the controllers. The controllers manage the control actions. Feedback to the controller is used to observe the controlled process and make updates to the control actions.

In fully autonomous vehicles controllers consisting of algorithms are responsible for generating and controlling majority vehicle Systems without human involvement. Similarly, the human driver are controller of in traditional conventional vehicle as well as cyclists and pedestrians. Introduction of fully autonomous vehicles in the current traffic system increases the interaction between humans and machines controller hence further increasing the complexity of traffic systems.

Human behavior and psychology makes human controller complex. Literature provides a mental model for human controllers as shown in Figure 3.3 below.

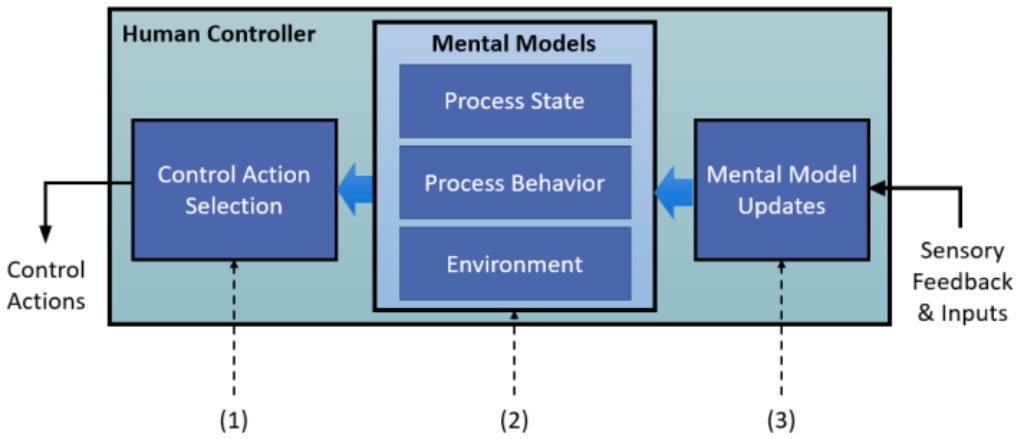

Figure 3.3: Engineering for Humans (France,2017)

To model a traffic system control structure, knowledge of the functional architecture of autonomous and conventional vehicle is vital. There is not much literature available publicly on fully autonomous vehicle functional architecture however, a number of STPA control structures for semi with a human driver and fully autonomous vehicle can be found in literature.Examples of such control structures are highlighted in Figures 3.3 and Figure 3.4.

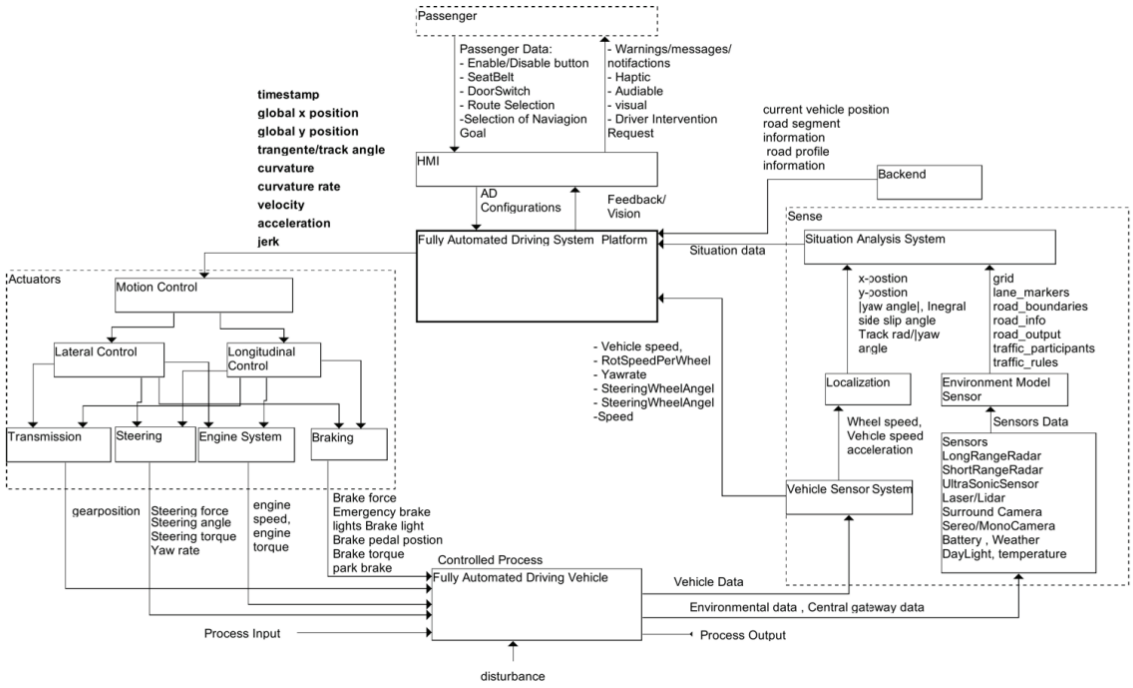

Figure 3.4: The control structure diagram of the fully automated driving vehicle (Abdulkhaleq et al., 2017)

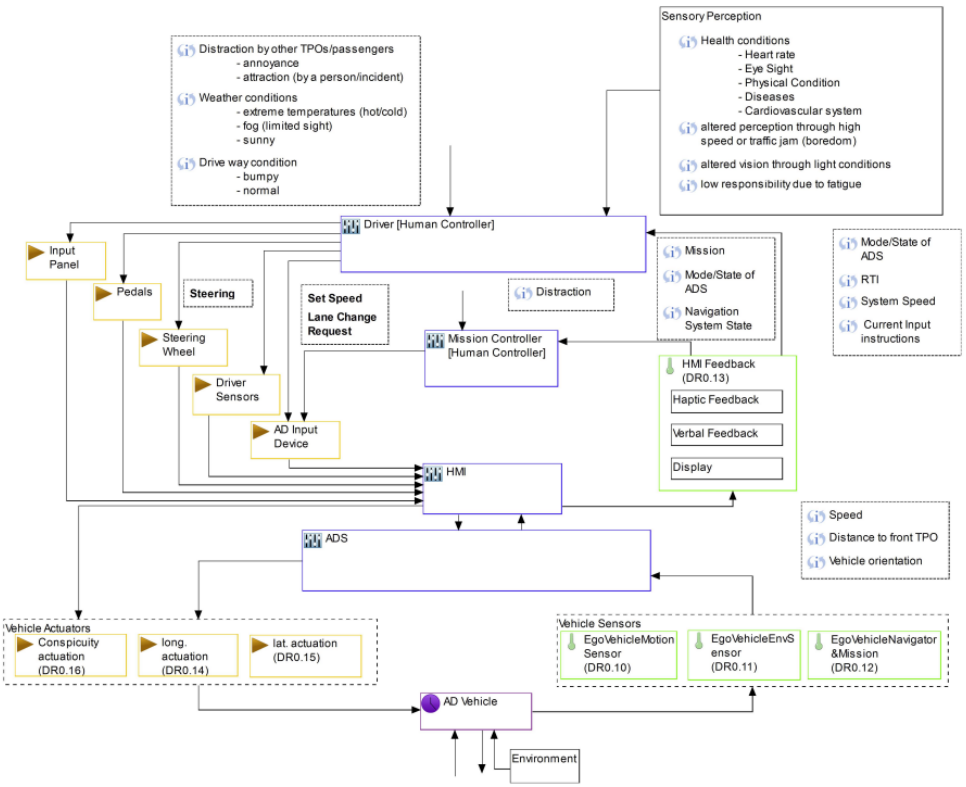

Figure 3.5: Control Structure for the human driver (Abdulkhaleq et al., 2017)

Using examples from literature a traffic system control structure can be modeled

by combing all traffic participant control structures to each other.Linking them by their interactive relationships. To reduce on complexity the traffic control structure can be generic, with mostly high level detailing.

3.1.3 Unsafe control actions

The unsafe control actions that each component may implement are identified in this step, which aids in refining the safety restrictions and requirements for the system. It will outline the justifications for these risky control measures. The control actions that potentially result in accidents that are used to define UCAs. As a result, this analysis uses the control diagram to consider four control actions for the study.

The following are the reasons that a control action can lead to an unsafe scenario.

1) If a control action is not given.
2) If a risky controlling action is given.
3) If a controlling action is provided at the incorrect time (too early or too late)
4) If a control action is applied for a prolonged period or is terminated too early.

3.1.4 Identifying loss scenarios

This is the final step of the STPA process and it involves systematically exploring various possible causes leading to unsafe control actions. To have a systematic process of hazard casual factor identification adjacency matrices were used as shown figure 3.6 below. Where H=hazard, UCA= unsafe control action and CF=casual factor. In the matrix 0 means there is no relation and 1 means there is a relationship. Using adjacency matrices casual factors can be identified at high to low levels in a structured way.

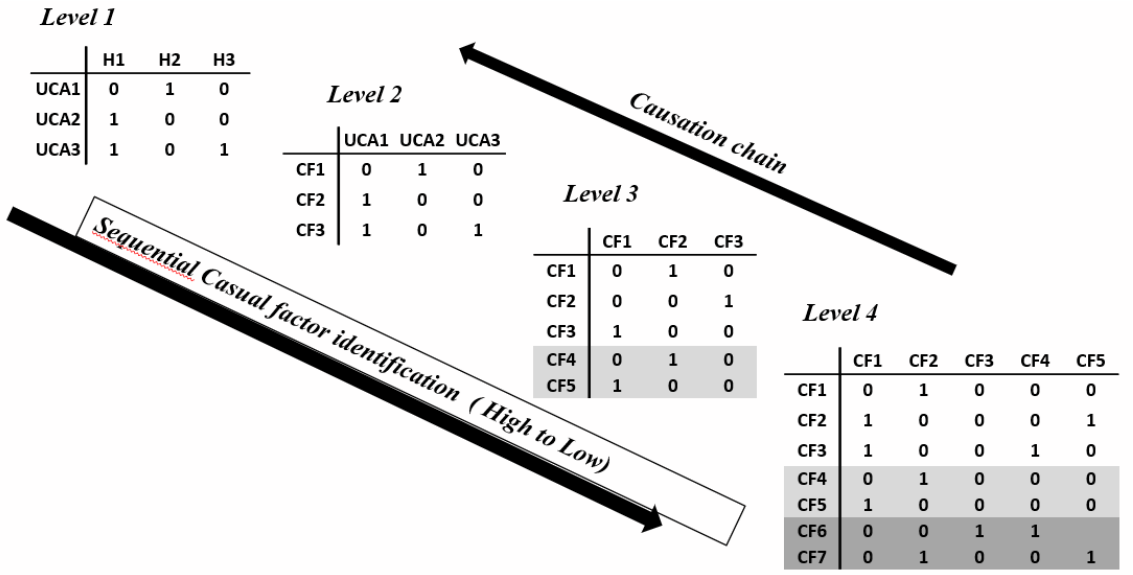

Figure 3.6: Adjacency Matrix for a Structured Casual factor Identification Process

The identified causes are used to formulate the safety requirements for each specific

causes/loss scenario.This part of STPA evolves brainstorming and relies heavily on the expert domain and works best if performed in a group of people with adequate knowledge of the system under consideration. However, an exposure to knowledge on traffic systems and autonomous driving helps the process.

3.1.5 Complex network theory for casual factor evaluation

Since the STPA method is a qualitative method mainly used for identifying hazard casual factors. Using STPA alone it is difficult to valuate hazard casual factors. In view of this, complex network theory is a vital tool for analyzing complex systems to quantitatively evaluate hazard causal factors. In this thesis a Hazard Causation Network was generated and topological indices were applied. Networkx in python was used for this process.

3.1.5.1 Construction of Hazard Causation Network (HCN)

The Hazard Causation Network is constructed through an adjacency matrix defined below.

$$AM_{ij} = \begin{cases} 1\,,\ (i,\ j) \in R \\ 0\,,\ (i,\ j) \notin R \end{cases}$$

Where, i and j represents the hazard causes respectively, and the cause-effect relationship between them . R represents either a true or false relationship. The above equation shows that there is a directed edge representing cause-effect between the two causes. In this thesis the adjacency matrix was generation during the last step of the STPA process as described in sub section 3.1.4. The generated matrix was later exported to python networkx to generated the Hazard causation Network as illustrated in figure 3.7 below.

	H1	UCA1	UCA2	CF1	CF2	CF3
H1	0	0	0	0	0	0
UCA1	1	0	0	0	0	0
UCA2	1	0	0	0	0	0
CF1	0	1	2	0	0	0
CF2	0	0	1	0	0	0
CF3	0	0	0	1	1	0

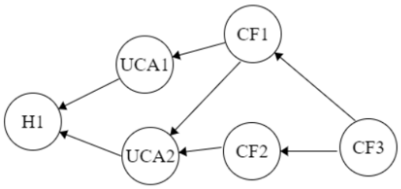

Figure 3.7: Adjacency Matrix to Hazardous Factor Network

3.1.5.2 Causal factor evaluation

The HCN shows the topological structure of hazard causation factors from the perspective of casual relationships hence, providing a model foundation for evaluation. A topological index inbetwenness was proposed in this thesis to measure the influence of each causal factor based on it's connectivity within the HCN.Networkx was used to find the inbetweeness values for all causal factor for comparison. Lastly to

show the chain of connectivity between causal factors to each other, causal factor to unsafe control action and finally to an identified hazard, hazard-casual-factor traceability networks were extracted from the HCN using netwokx. The final evaluation is illustrated in figure below.

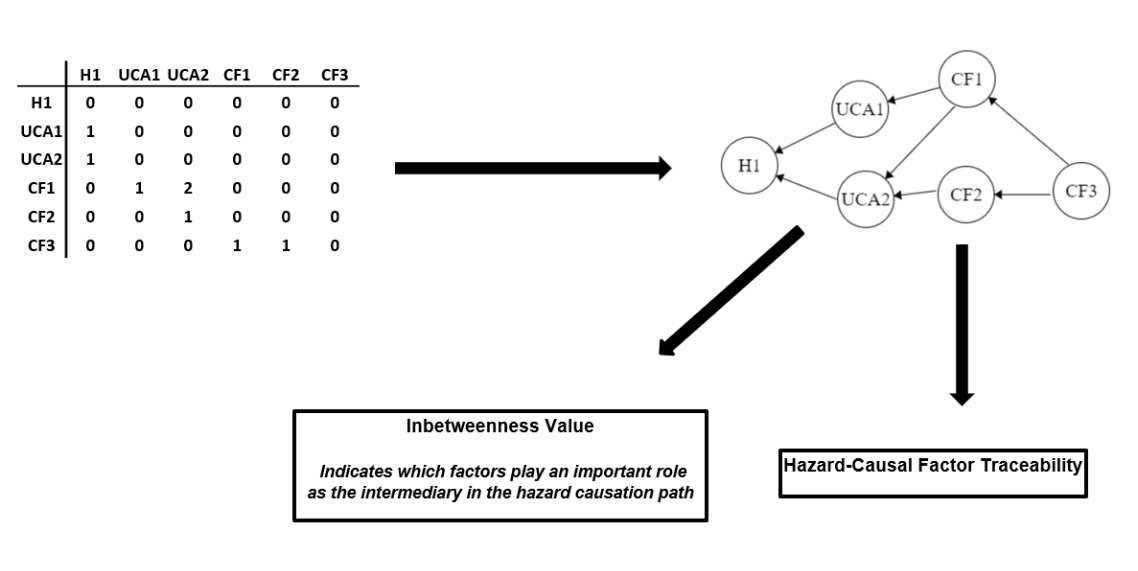

Figure 3.8: Causal factor evaluation process

4

STPA analysis and results

The aim of this thesis is to provide a pilot application of STPA for evaluating the safety of Autonomous Vehicles in different traffic scenarios within the current traffic system .Therefore, STPA was applied to a scenario-based safety analysis of an autonomous vehicle at an unsignalled 4-leg intersection. The output of this thesis provides a framework for applying STPA in traffic systems safety analysis, generating knowledge and practices for further development. This chapter deals with the application of STPA as outlined in chapter 3.

4.1 Traffic scenario description

This thesis investigates an unsignalized intersection where level 4 autonomous vehicles have been introduced, and safety based on the STPA criteria is studied.

In road traffic systems, intersections contribute to high levels of traffic accident cases. In general, the 4-leg corner of two 2-lane, 2-way roadways have 32 conflict points, including eight merging points, eight diverging points, and 16 crossing conflict points, as illustrated in Figure 4.1 below.

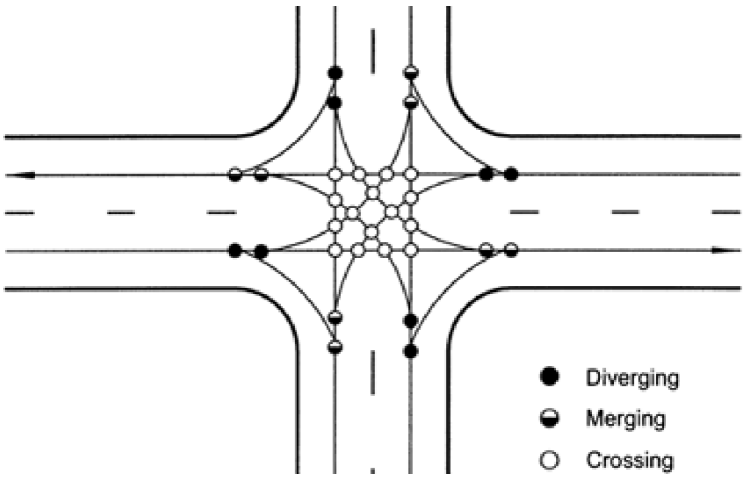

Figure 4.1: Conflict Points at an intersection (FHWA 2004)

Studies have shown that most accidents at intersections are due to unsafe interactions between traffic system participants. Statistical analyses of the causes of accidents at intersections have shown that 89 % of them are due to driver error. The most common mistakes are perception failures, situation misunderstanding,

and wrong decisions. The causes of these errors are usually inadequate surveillance of surroundings, false assumption of other drivers' actions (poor communication of intent), turning with an obstructed view, illegal maneuvers, internal distraction, and misjudgment of other drivers' speed.

4.2 Purpose of analysis

The primary step in the STPA process involves defining the purpose of the analysis. It includes identifying losses unacceptable to stakeholders, hazards that may lead to the highlighted losses, and defining system-level constraints. Foremost, the system to be analyzed must be identified, its boundaries and components defined. The system also needs a well-defined goal or objective. The components in a traffic system are also systems, making it a system of systems. A scenario-based STPA safety analysis is applied in this thesis, where the system goal or objective is the safe navigation of a 4-leg intersection with the introduction of fully autonomous vehicles.

4.2.1 System definition

The traffic system was identified as constituting the road authorities and car manufacturers, fully autonomous vehicles (level 4), conventional vehicles, cyclists, pedestrians and infrastructure inside the domain of control, whilst weather condition are outside the domain of control but still included due to interaction with all traffic participants. The traffic system is illustrated as shown in figure 4.2 below.

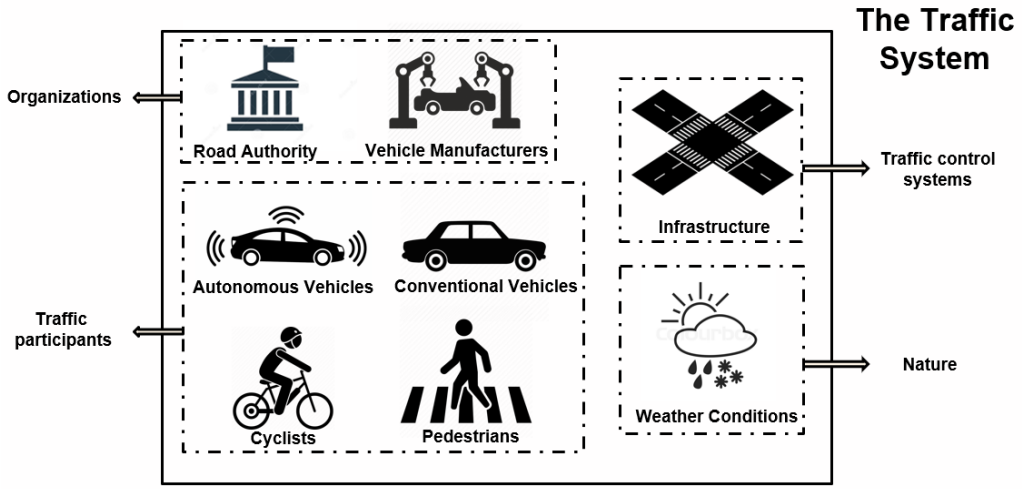

Figure 4.2: The traffic system

%
The road authority has significant high-level control over human behavior and vehicle manufacturers through regulations and setting functional standards. Hence, it is important to include the road authority in the traffic system under analysis. Similarly, vehicle manufacturers have high-level control over the performance capabilities of both autonomous and conventional vehicles in the traffic systems. However, they

Table 4.1: Different Systems in Traffic environment.

Traffic organizational systems	**Road authority**	Promote road safety
		Research road accidents and safety
		Driver testing and licensing
		Vehicle related safe driving standards
	Vehicle manufacturers	Design and Development
		Performance testing and validation
		Manufacturing
		Quality assurance and quality systems
Motorized traffic participants (systems)	**Autonomous Vehicles**	Perception (hardware)
		Decision (software)
		Identification (software)
		Execution (vehicle actuators) standards
	Traditional conventional Vehicles	Perception (human)
		Decision (human)
		Identification (human)
		Execution (vehicle actuators) systems
Non-motorized traffic participants (systems)	**Cyclists**	Perception (human)
		Decision (human)
		Identification (human)
		Execution (bicycle actuators) standards
	Pedestrians	Perception (human)
		Decision (human)
		Identification (human)
		Execution (body actuators) systems
Traffic control systems	**Infrastructure**	Road signs
		Streetlights
		Lane markings
		Pavement type
		Lane width
		Speed limit
Natural elements	**Weather conditions**	Extreme temperature
		Fog
		Sunny
		Rainy
		Snow

have no control over human driver behavior. Including the vehicle manufacturers as part of the traffic system is also vital. Motorized, non-motorized traffic participants and traffic control systems are the main constituents of the traffic system. The natural elements of the environment, such as weather conditions (rain, snow, extreme temperature)s are not inside the domain of control. However, their impact on the traffic system is significant. Therefore, it is important to include weather conditions in the analysis, along with the traffic system. The roles of each member in the traffic system is highiligted in table 4.1.

4.2.2 Losses

After defining the traffic system and scenario for analysis, losses are defined based on what is deemed unacceptable by stakeholders, the loss of something of value.

Stakeholders, in this case, are the road authority, vehicle manufacturers, and traffic users. For the system, as defined in this study, the losses were identified and highlighted in Table 4.2.

Table 4.2: Identified losses (Extracted from A-STPA).

No	Title	Description
1	Loss of life or serious injuries	Loss of life or injury due to serious collision
2	Damage to property	Damage to property due to collision
3	Loss of travel time	Time loss or mission loss
4	Road authority loss of reputation	unreliable traffic system may lead to the road authority losing their reputation
5	vehicle manufacturer loss of reputation	unreliable vehicles in a traffic system may lead to the vehicle manufacturer losing their reputation

4.2.3 Hazards

The next step in the STPA process is to identify hazards. Hazards, in this case, are system states or sets of conditions that lead to the above-highlighted losses in a particular worse-case environment condition. The hazards are linked to one or more losses identified in table 4.2.A total of 6 hazards were looked into when performing the STPA analysis. There can be more hazards considering the lossess, however there were the most common hazards when traffic environemnt is taken in consideration. The identified hazards are highlighted in table 4.3 below.

Table 4.3: Identified Hazards (Extracted from A-STPA).

No	Title	Description
1	Illegal driving manoeuvres	Traffic participant violates stated rules for flow of traffic, including right of way
2	Poor communication of intent	Communication of intent depends on cues in road participant's behaviour such as eye contact, posture, gesture and external vehicle interface or signals.
3	Misunderstanding, wrong judgement and decision	Wrong perception and decisions at intersection can lead to high probability of an accident event
4	Traffic participants unable to keep safe distance with surrounding infrastructure	If traffic participants cannot keep a safe distance with surrounding infrastructure, it may lead to an accident event.
5	Traffic participants unable to maintain safe distance with each other	If traffic participants cannot keep safe distance with among each other, it may lead to an accident event.
6	Unreliable autonomous driving functionalities	If autonomous functions are not reliable, it may lead to an accident event.

4.2.4 System-level constraints

System-level constraints are system conditions that must be satisfied to prevent the traffic system from being in a state of hazard. System-level constraints are simply an invert of conditions that result in losses. All the traffic participants have their own constraints put in place so that any potential loss scenario can be avoided.

4.3 Control Structure

As discussed in the previous chapter, the second stage in the STPA process is modeling the traffic system control structure. The control structure is composed of control loops and feedbacks forming a system model. The controllers enforce constraints on the controlled process by providing control actions. The controller receives feedback from the controlled process of which the controller uses for observation and updating decision-making. The control structure is usually hierarchical, where some controllers exert constraints on lower-level controllers.

The most immense challenge in hazard analysis of traffic systems is handling their complexities. For example, a highly complex traffic system has multiple controllers and multiple controlled processes. The traffic system under analysis comprises the following systems: the road authority, vehicle manufacturers, autonomous vehicles, conventional vehicles, cyclists, pedestrians, and traffic infrastructure. Control structures use abstraction in several ways to help manage these complexities. To simplify the complexities of a traffic system when modeling the control structure, firstly, control structures will have to be modeled for individual system in the traffic system.

Road authorities and Vehicle manufacturers
The road authorities are a very active stakeholder in the traffic system. Hence, are one of the main controllers in the system. They establish all regulations and laws that form the basis on which the traffic system functions. The organizational structure of the road authority can be complex and differ from one municipality to another. In this research, no attempts are made to define the internal components that make up the organization system of the road authority. Therefore, the control structure for the road authority system is defined at the highest system level.
The performance of autonomous and conventional vehicles in the traffic system depends on the design and developments carried out by individual vehicle manufacturers under standards and regulations set forth by the road authority. Therefore, vehicle manufacturers are also an important component of the overall traffic system. In general, vehicle manufacturing is dominated by the private sector; hence the inner working of vehicle manufacturers and their organizations can differ from company to company. Just as in the case of road authorities, the vehicle manufacturer control structure is defined at a high system level.

Autonomous vehicles (level 4 automation)
Fully autonomous vehicles are highly reliant on software, and vehicular networks which enable the vehicle negotiate an intersection. This makes defining the con-

trol structure for autonomous vehicles difficult. One way of developing the control structure for Avs is by using the functional architectures of fully automated driving vehicles as published in the existing literature on autonomous driving. In this thesis, the control structure for autonomous vehicles is obtained from (Abdulkhaleq et al., 2017) as highlighted in the methodology and simplified as follows.

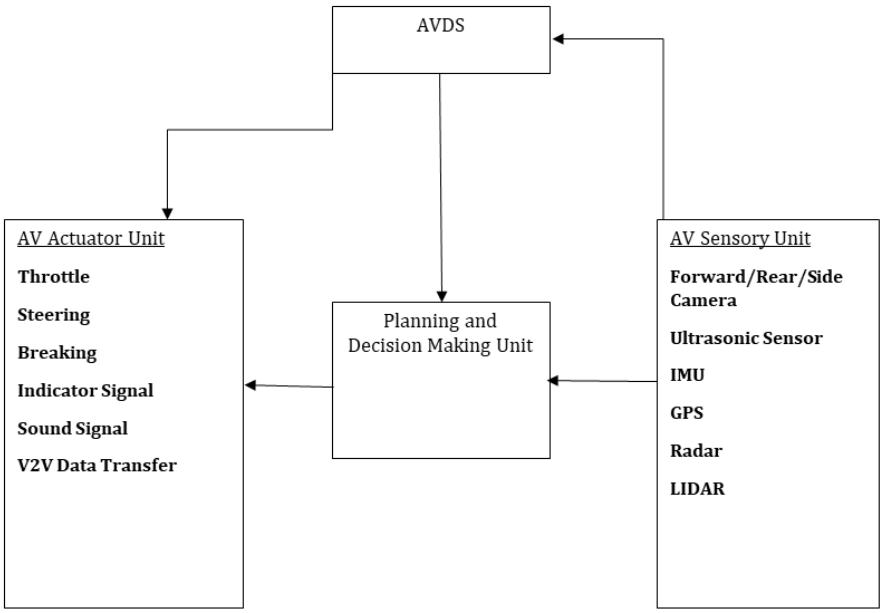

Figure 4.3: Autonomous vehicle control structure

4.3.1 Conventional vehicles, cyclists, and pedestrians

The control structure for conventional vehicles (CVs) developed resembles that of AVs, except sensors and the automated driving system are replaced by human sensory and human driving. Just as in the conventional vehicle control structure, the controller for cyclists and pedestrians as traffic participants is the human being. The difference between conventional vehicles, cyclists, and pedestrians is the availability of actuators and the controlled process.

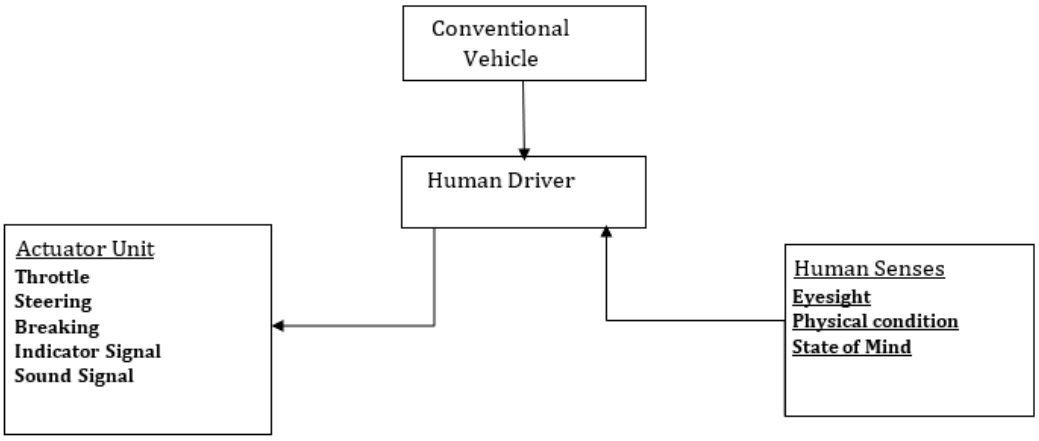

Figure 4.4: conventional car control structure

4.3.2 Traffic system control structure

The traffic system control structure combines the above-mentioned systems connected by their interactions. A high-level traffic system control structure was modeled at an abstract level, consisting of function blocks relevant to negotiating an intersection for each system within the overall traffic system. The control structure is shown in Figure 4.4 below. Functional blocks represented by stakeholders, i.e., road authorities and vehicle manufacturers, are illustrated as controllers only, whilst systems representing traffic participants are divided into the following functional architectural parts.

Controllers are the planning and decision-making units, consists of a control algorithm and internal belief. In a fully AV, the controllers are the AV driving system platforms for planning and decision. For CVs, cyclists and pedestrians, controllers are human beings. Actuators which deal with motion control have the task of performing lateral and longitudinal motions. These include steering, braking, engine control, and transmission. For cyclists, the two latter are omitted; for pedestrians, all actuators are part of the human body. For AVs, there is also the possibility for vehicle-to-vehicle real-time data sharing of the condition-controlled process and planning decisions from the controllers. The controlled process is another function block. It is observed that controlled process for vehicles, both autonomous and conventional, is the vehicle itself. In contrast, for cyclists, it is the bicycle, and for pedestrians, it is the body. Sensors deal with perception and observation of surroundings and object recognition of traffic signs, where road and lane detection are vital for all traffic participants. It includes the ability to recognize the presence of other traffic participants and understand their intent through communication. AVs use various hardware and software components for perception ability, such as sensors, radars, cameras, Light Detection and Ranging (LiDAR) systems, and Global Positioning System (GPS) for localization and mapping. For CV, perception is solely dependent on human capabilities.

After identifying each system that is a part of the overall traffic system, the next step is establishing their interactions. The traffic system is both hierarchical and non-hierarchical (collaborative). The interactions between road authorities and vehicle manufacturers are hierarchical and composed of a feedback control loop. Thus, the road authority enacts ISO standards such as ISO 26262, which go down as control actions and the car manufacturers give back feedback in the form of adherence reports which includes vehicle performance information. The relationship between vehicle manufacturers' Avs and CVs in a traffic system is also hierarchical. As designers and developers, vehicle manufacturers perform performance tests or observations on their vehicles in a traffic system to extract performance indicators. For the systems considering traffic participants, the interactions amongst each other are non-hierarchical, and in a traffic system (intersection scenario), communication of intent is the main and only interaction described. As for the road authority to the traffic participants, the relationship is hierarchical, with the road authority enacting traffic laws and public options as feedback. The system connection between the road and infrastructure is also hierarchical. The road authority's control action is to maintain and modify the infrastructure.

The weather conditions and infrastructure constitute the measured operation design domain. There is no control over weather patterns; hence, for safety, the road authority must evaluate the performance of traffic participants in relation to the operational design domain, make modifications to the infrastructure, and update ISO standards and traffic laws. Weather condition data is also collected by vehicle manufacturers and traffic participants for their decision-making relating to the operational design domain.

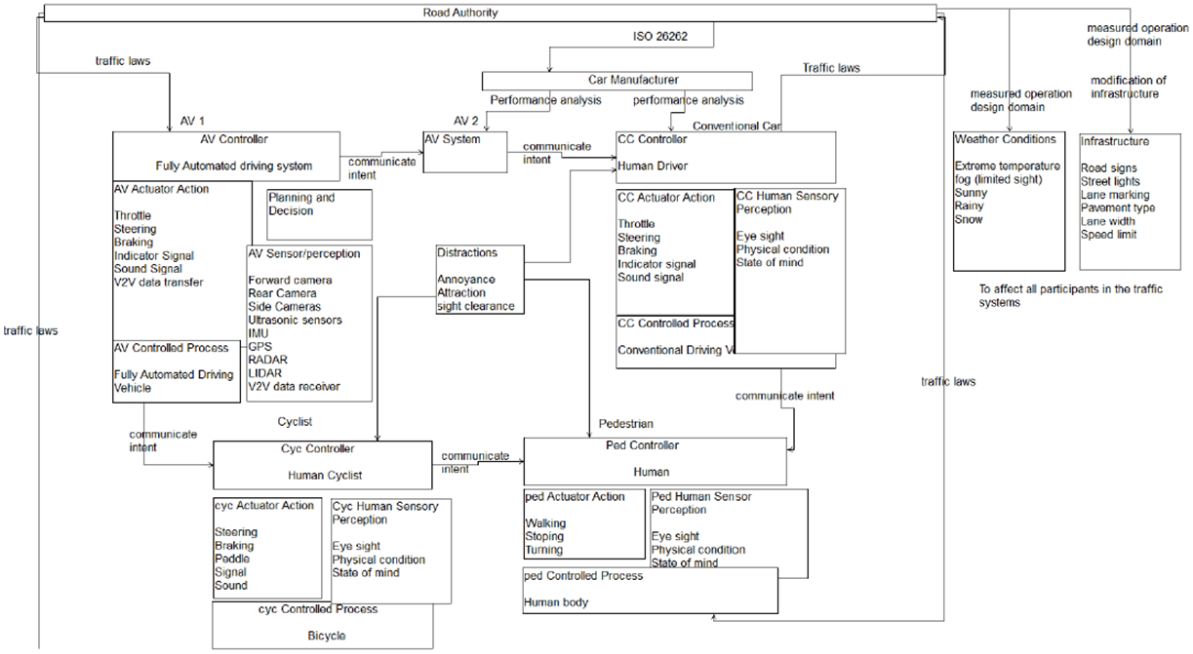

Figure 4.5: Complete control structure

4.3.3 Unsafe control actions

The next step after drawing the control structure in the STPA process is identifying the UCAs. The UCAs, as mentioned in the previous chapter, are control actions that in the worst case environment and in a particular context will will lead to a hazard.

For a traffic system scenario of an AV at an unsignalized intersection, four control actions were identified: Yielding when not having the right of way by all traffic participants, maneuvering of the intersection by all traffic participants, communication of intent by all traffic participants, maintenance, and modification of infrastructure components by the road authority.

Table 4.4: Identified UCA (Extract from A-STPA).

Control Action	**Not Given**	**Given Incorrectly**	**Wrong Timing or Order**	**Stopped too soon or applied too long**
Yield	AV does not yield when not having right of way. [1, 5, 6]	Av yields too early or too late before or after the stop line. [1, 6]	AV yields abruptly	
Turning Manoeuvre	AV cannot perform a turning manoeuvre at the intersection [6]	AV perform incorrect turning manoeuvre at the intersection [1, 4, 5, 6]		
Communication of Intent	Av fails to communicate intent [2, 3]	AV communicate misleading intent [2, 3]	AV delays to communicate intent [2, 3]	
	AV cannot receive communication of intent from other traffic participants [2, 3]	AV receives misleading communication of intent from other traffic participants [2, 3]	Communication of intent from other traffic participant id delayed [2, 3]	

4.3.4 Identifying loss scenarios

After identifying the UCAs, the next step in the STPA process is the identification of loss scenarios. Loss scenarios explains the causes of UCAs and consequential hazardous scenarios.

Causal factors were identified by answering the following questions:

a) Why would UCAs occur? This deals with sensors, feedback processes, the control process, and the control algorithm.
b) Why would control actions fail to be executed properly or not executed entirely? This deals with actuators and other components involved in the system process during the execution phase of the control action.

73 Causal factors to the earlier identified UCAs were identified by analyzing the interactions in the control structure. Due to limited space the full list of identified causal factors is attached in the appendix. The tables below are a sample of unsafe control actions and their respective causal factors as extracted from the excel adjacency matrix.

			UCA1	UCA2	UCA5
			AV does not yield when not given right of way	AV Yields too early or too late before or after stop line	AV incorrectly manoeuvres intersection
Car Manufacturer	CF5	Inadequate vehicle automation design by manufacturer	1	1	1
Infrastructure	CF7	Missing road signs at intersection	1	0	0
	CF8	No road lane markings at intersection	0	1	1
	CF9	Intersection and speed signs too close to intersection	0	1	0
	CF10	intersection and speed signs too far away from intersection	0	1	0
	CF11	Road lane markings are not visible, covered with dirt	0	0	1
	CF13	Obstructed vision (short sight distance to intersection)	1	1	0
Weather	CF19	Slippery surface due to snow or rain	0	1	0
	CF20	Rainy or Foggy weather limiting sight	1	1	1
	CF21	Sunny (glaring light on sensors)	1	1	0
Autonomous Vehicle (1)	CF22	AV has misleading perception of road signs at the intersection approach	1	1	0
	CF23	AV planning and decision error	0	1	1
	CF24	AV has wrong speed perception of other motorized traffic participants	1	0	0
	CF29	AV can not detect lane markings	0	0	1
	CF30	AV has wrong perception of cyclist and pedestrian intent	1	0	0

		Level 3	CF29	CF30	CF31	CF32	CF33	CF34	CF36	CF38
			AV can not detect lane markings	AV has wrong perception of cyclist and pedestrian intent	AV cannot understand vehicle communication signals through use of turn signals, brake lights, hazard lights, headlights, and horn	AV has wrong perception of intent of other motorised traffic participants	AV has wrong locality perception (distance between AV and intersection)	AV unable to transfer real-time vehicle information to AV2	AV is unable to give vehicle communication signals (turn signals, brake lights, hazard lights, headlights, and horn)	AV2 sends misleading communication of intent
Weather	CF19	Slippery surface due to snow or rain	0	0	0	0	0	0	0	0
	CF20	Rainy or Foggy weather limiting sight	0	1	0	1	0	0	0	0
	CF21	Sunny (glaring light on sensors)	0	1	0	1	0	0	0	0
Autonomous Vehicle (1)	CF22	AV has wrong perception of road signs at the intersection approach	0	0	0	0	0	0	0	1
	CF23	AV planning and decision error	0	0	0	0	0	0	0	0
	CF72	Actuator component failure	0	0	0	0	0	1	1	1
	CF73	Perception component failure	1	1	1	1	1	0	0	0
	CF24	AV has wrong speed perception of other motorized traffic participants	0	0	0	0	0	0	0	1
	CF25	AV lacks ability to transfer non-verbal communication cues	0	0	0	0	0	0	0	0
	CF26	AV lacks ability to receive non-verbal communication cues	0	1	0	0	0	0	0	0
	CF27	AV information system security breach	0	0	0	0	0	0	0	0
	CF28	AV non-verbal communication cues ability is not effective	0	1	0	0	0	0	0	0
	CF29	AV can not detect lane markings	0	0	0	0	0	0	0	0
	CF30	AV has wrong perception of cyclist and pedestrian intent	0	0	0	0	0	0	0	1
	CF31	AV cannot understand vehicle communication signals through use of turn signals, brake lights, hazard lights, headlights, and horn	0	0	0	1	0	0	0	0
	CF32	AV has wrong perception of intent of other motorised traffic participants	0	0	0	0	0	0	0	1

4.3.5 Hazard Causation Network

The hazard causation network HCN was generated using networkx. Figure 4.6 below illustrates the HCN consisting of 84 nodes of CFs and UCAs, and 198 edges.The figure shows how highly connected the causal factors are to each other and to unsafe control actions.

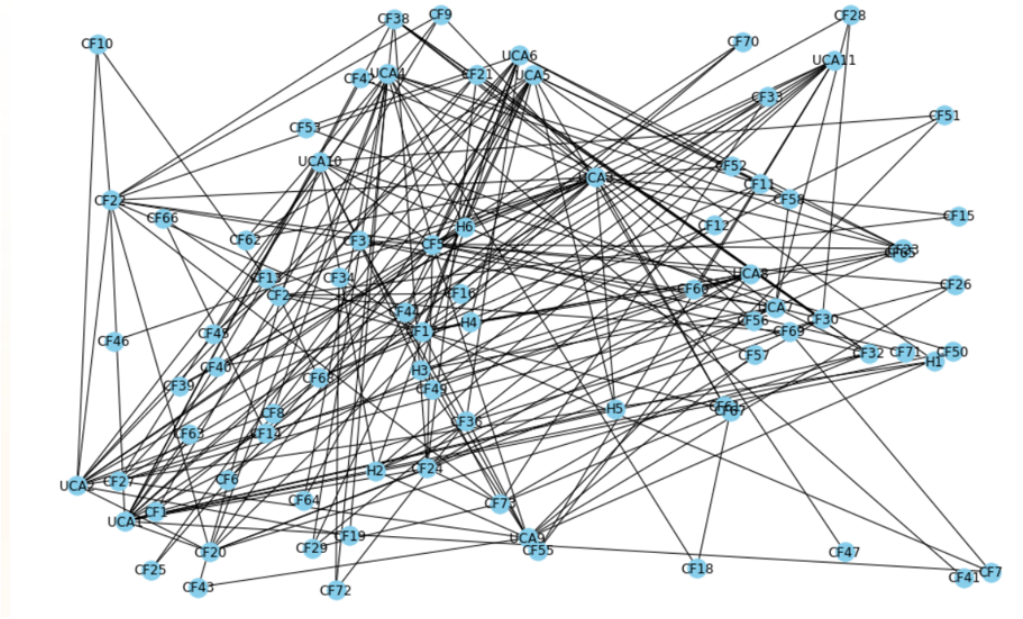

Figure 4.6: Hazard Causation Network

4.3.6 hazard causal factor evaluation and treceability

The values of inbetweeness nodes were calculated using networkx and shown in figure 4.7 below. It can be seen that UCA 5 has a higher inbetweenness value than the

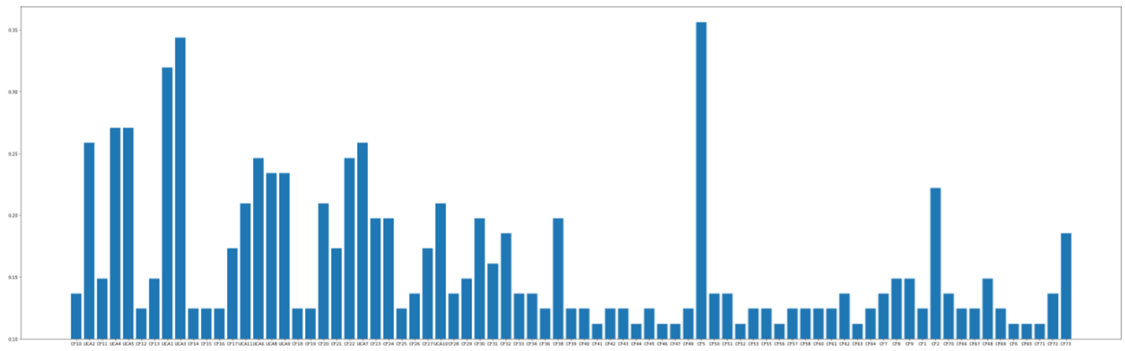

Figure 4.7: Betweenes value for UCAs and CFs

other UCAs because it leads to the highest number of hazards. CF5 (inadequate road infrastructure maintenance and management plan by road authority) has the highest inbetweenness values among causal factors. Hence, maintaining infrastructure within the operation design domain is vital for the safety of fully AVs in a

traffic system. If the causal factors with high inbetweeness values can be controlled, a significant portion of causal paths leading to hazards can be blocked.

Causal factor traceability is also important in identifying causal paths leading to hazards. Casual paths were identified for all paths and an illustration of causal path for Hazard H1 (Illegal driving manoeuvres) is show in figure below.

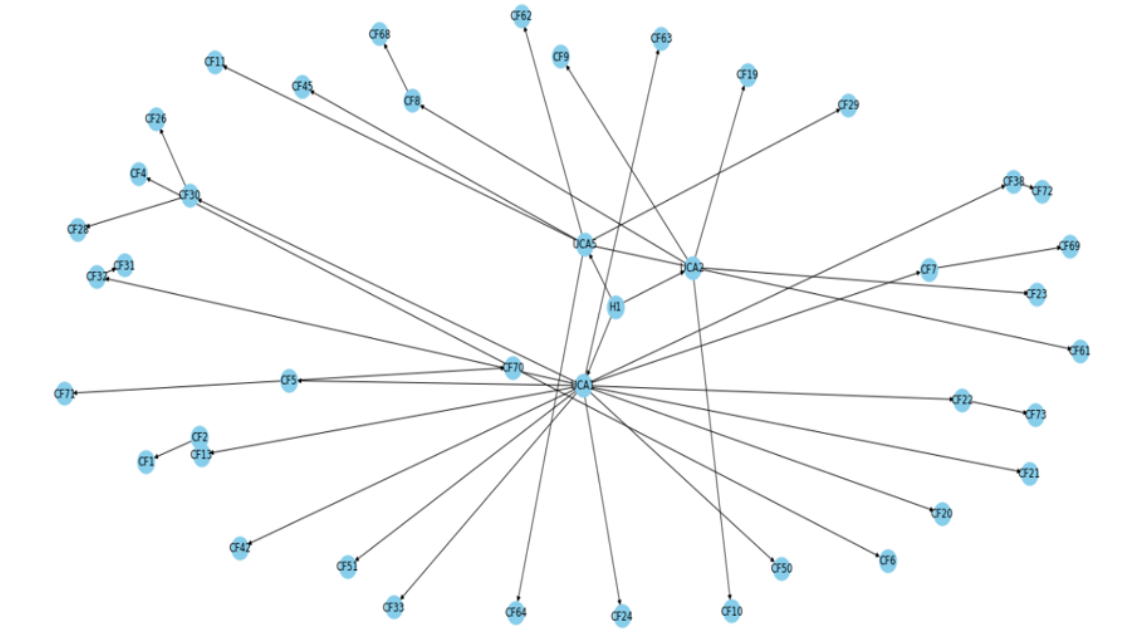

Figure 4.8: Hazard causal path for H1

5

Discussion

5.1 Reflection on STPA approach

The first step in the STPA hazard analysis involves defining the system under consideration. This process proved not an easy task when the system under analysis is itself comprised of subsystems interacting with each other. A traffic system is one such system. The introduction of an autonomous vehicle into the current traffic system increases the complexity of the traffic system and this complicates the interactions with all other traffic participants and the overall functionality of infrastructure and traffic regulations. This significantly changes the traditional way a traffic system operates. Therefore, for autonomous vehicles to operate safely in the current traffic system several organizations such as government regulators, local municipalities, and vehicle manufacturers together with all other traffic participants need to actively evolve and adapt to the changes arising from the use of fully autonomous vehicles. In the STPA contest, this means describing a traffic system that takes all participants into consideration a difficult task. The risk is overcomplicating or oversimplifying the system and not achieving a comprehensive hazard analysis.

The second step in the STPA process is modelling the control structure. The control structure is highly significant in the STPA process because it's from the modelled control structure that controls, and feedback interactions are determined. It's from these interactions unsafe control actions that lead to hazardous scenarios being identified. Hence modelling the control structure has a significant effect on the overall safety analysis. As earlier stated, the introduction of fully autonomous vehicles into the current traffic system increases the complexity of such systems. For the traffic system as a System of systems, each traffic participant, organizations that regulate operational design domains, and infrastructure and vehicle manufacturers that ensure developed vehicles conform to specifications has to be represented in the control structure. This makes modelling the control structure a complex and difficult task. To get around this problem a lot of generalisation and simplifications have to be made. This overall affect the comprehensiveness of the hazard analysis. The other difficulty in modeling the control structure is the lack of public information on the organisational structures and functional architectural models for each traffic participants. In this thesis, the control structure was modelled based on the information that could be found in literature.

The last two steps in the STPA process involve identifying unsafe control actions and

their respective causal factors. To identify unsafe control actions first, the system control actions had to be identified. It was observed during the analysis that when using STPA for hazard analysis in a traffic system the scenario-based approach was the best way to identify system control actions rather than generalising. Due to the complexities of the traffic system as a system of systems and the further complication of having fully autonomous vehicles interacting with other traffic participants, the identification of the causal factors can easily become a messy process with several interconnections. The STPA process as outlined in several pieces of literature on its own does not provide in detail an organised and structured way of casual factor identification, especially for highly complex systems. To solve this problem a system of adjacent matrices was used. This proved effective in structured the casual factor identification process.

5.2 Reflection on the hazard analysis results

The final result of the STPA analysis was a list of identified hazard causal factors. There was always a concern during the analysis on the maturity and significance of the final results. It's highlighted in literature that one of the problems analysts face when applying STPA is the simplicity of results obtained. A number of identified causal factors in this thesis could not have been identified without using the STPA framework however, a number of identified causal factors were already known.

The STPA tool proved to be a powerful tool for discovering the complexities of introducing a fully autonomous vehicle in the current traffic system. However the results did not have vast depth of maturity because a number of factors. Firstly there is a weak link between the identified causal factors and the unsafe control actions due to lack of quantitative classification of causal factors and unsafe control actions. Secondly the result from the STPA analysis greatly depend on the knowledge about the System and the components. The STPA method depends on the amount of information about the System under consideration. The STPA process is also time consuming. To have an in-depth analysis a lot of time need to be spend during the brainstorming process which was not the case in this thesis research

The results from complex network analysis help quantify the STPA results to a degree. Using inbetweeness values to determine the most influential hazard causal factors was an easy evaluation tool. The results showed that the more general casual factors tend to have high inbetweeness values. Hence in depth detailing during casual factor identification can produce more mature results.

Combing STPA with complex network theory proved to be quite usufully for a systematically hazard causal identification process, expecially for complex networks such as a fully autonomous vehicle in the current traffic system.

6
Conclusion and Limitations

This thesis presents a pilot application of STPA for evaluating the safety of introducing fully autonomous vehicle in the current traffic system. STPA was applied to scenario-based safety analysis of AVs at an unsignalled 4-leg intersection. The results provide a framework and demonstration of applying STPA in traffic systems safety analysis. The results of this thesis have shown that STPA is an efficient and effective hazard analysis method, suitable for assessing the safety of complex systems such as a traffic system. Using STPA 73 hazard causation factors were identified. Complex network theory was also applied to further evaluated identified causal factors. This gave the results a more industry worthy application.

The modeled control structure showed that the road authority are important and have high level control in the traffic system hence,hazard causal factor originating from actions of the road authority had more relevance. Finally the results showed that the STPA results would more significant if there was a "probabilistic link" between the Unsafe Control Action (UCA) and the causal factors. Adding these probabilistic links need significant time speed on the getting an insight of the operations of the traffic system and there maybe a need to collect substantial amounts of data.

Finally it is recommended that further research should be done combining STPA with other quantitative systems analysis tools such as system dynamic models which can potentially give us cause and effect relationships between the factors in numerical terms. Further,it was observed that STPA is very strong when finding causal factors for identified hazards. So combining it with FMEA, which takes into account the architecture of components, will enhance the risk assessment process.Moreover, there is alot of research going on with STPA extensions. One of which is STPA - Sec which combines safety with security.Apart from A-STPA and XSTAMPP, more tools can be developed which provide more functionality and better interface for risk assesments.

As with any research, there are limitations to this thesis. Two main limitations are worthy noting. Firstly, STPA is a time consuming process, and this thesis did not accommodate both budget and time constraints. Secondly, detailed information on AV functional architectural vital for control structure modelling was not readily available and attempts to obtain such data from private companies in the automobile industry were futile.

References

Abdulkhaleq, A., Lammering, D., Wagner, S., Röder, J., Balbierer, N., Ramsauer, L., ... Boehmert, H. (2017). A systematic approach based on stpa for developing a dependable architecture for fully automated driving vehicles. *Procedia Engineering*, *179*, 41–51.

Drive me: Volvo cars' approach to autonomous driving. (2016). Retrieved from `https://www.media.volvocars.com/global/en-gb/media/videos/189739/drive-me-volvo-cars-approach-to-autonomous-driving`

Hughes, B., Newstead, S., Anund, A., Shu, C., & Falkmer, T. (2015). A review of models relevant to road safety. *Accident Analysis & Prevention*, *74*, 250–270.

ISO, I. (2011). 26262: Road vehicles-functional safety. *International Standard ISO/FDIS*, *26262*.

Koopman, P., & Wagner, M. (2016). Challenges in autonomous vehicle testing and validation. *SAE International Journal of Transportation Safety*, *4*(1), 15–24.

Laracy, J. R., & Leveson, N. G. (2007). Apply stamp to critical infrastructure protection. In *2007 ieee conference on technologies for homeland security* (pp. 215–220).

Leveson, N. (1995). *Safeware: System safety and computers. addison wesley, reading.* United States of America Massachusetts.

Michon, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do? In *Human behavior and traffic safety* (pp. 485–524). Springer.

Nelson, P. S. (2008). A stamp analysis of the lex comair 5191 accident. *Master's thesis, Lund*.

NHTS. (2013). Preliminary statement of policy concerning automated vehicles. *Washington, DC*, *1*, 14.

Pereira, S. J., Lee, G., & Howard, J. (2006). *A system-theoretic hazard analysis methodology for a non-advocate safety assessment of the ballistic missile defense system* (Tech. Rep.). Missile Defense Agency Washington DC.

SAE. (2014). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *SAE Standard J*, *3016*, 1–16.

Salmon, P. M., McClure, R., & Stanton, N. A. (2012). Road transport in drift? applying contemporary systems thinking to road safety. *Safety science*, *50*(9), 1829–1838.

Singh, S. (2015). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey* (Tech. Rep.).

Tingvall, C., & Haworth, N. (1999). Vision zero-an ethical approach to safety and

mobility. In *6th ite international conference road safety & traffic enforcement: Beyond 2000.*

Williams, A. (2015). Defining autonomy in systems: Challenges and solutions. *Issues for Defence Policymakers*, *27*.

Zhang, H., Li, W., & Qin, J. (2010). Model-based functional safety analysis method for automotive embedded system application. In *2010 international conference on intelligent control and information processing* (pp. 761–765).

Appendix

Table 6.1: List of Causal Factors identified

No	Causal Factors	Description
1	CF1	Poor financing of road authority activities which can further be linked to political or administrative issues in securing the required budget for road authority to build and maintain infrastructure necessary for AVs to function properly
2	CF2	Poor road infrastructure management and planning by road authority. This can be due to various reasons but majority could be because of administrative and lack of expertise in the transportation departments
3	CF3	Poor traffic law awareness by authority. Traffic laws are made based on incident reports and experts opinions. Laws that are made for conventional vehicles can not be enforced for AVs which can give rise to a hazardous situation
4	CF4	Road authority lacks expertise. Again qualified and technical people are sometimes not involved in the policy making process which can be unsafe in the long run
5	CF5	Inadequate vehicle automation design by manufacturer. This could be due to cutting costs or staying ahead in the market
6	CF6	Vehicle manufacturers do not regularly provide updated vehicle performance report to road authority
7	CF7	Missing road signs at intersection. Road signs are vital for AVs as their sensory system is dependent on these signs to act accordingly.
8	CF8	No road lane markings at intersection. It is important as lane keeping systems in AVs are dependent on the road markings and any missing marking could lead to an accident.
9	CF9	Intersection and speed signs too close to intersection. This will not give enough time for AV to reduce its speed when approaching the intersection
10	CF10	intersection and speed signs too far away from intersection. This is slow down the AV too far when approaching intersection causing disruption in the traffic flow
11	CF11	Road lane markings are not visible, covered with dirt. This can be attributed to road authority who is responsible for operation and maintenance of the roads.Since most AV system work identifying lane marking through sensors, any accumulated dirt can cause this functionality to not work resulting in hazardous situation
12	CF12	Confusing lane markings at the intersection (visible old marking, multiple lane markings)
13	CF13	Obstructed vision (short sight distance to intersection). This could be because of large vehicles or sometrimes trees outgrowth. AV systems needs clear vision to collect data and make decisions properly
14	CF14	Obstructed vision No sight clearance around intersection, tree too close to intersection
15	CF15	Obstructed vision, road side advertisement and notices

No	Causal Factors	Description
16	CF16	Obstructed vision No sight clearance around intersection, buildings too close to intersection
17	CF17	Faulty internet communication infrastructure (AV offline). GPS and Internet is a real time source of information for all AV systems and any disruption in this communication might be hazardous for this system
18	CF18	Incorrect roadway geometry at intersection (narrow roadway for AV design domain).If a road section has dimensions for which an AV system is not designed for can lead to a hazardous situation
19	CF19	Slippery surface due to snow or rain. AVs are tested in different weather conditions but extreme conditions are never the design conditions and AVs migh behave diffeently when external factors are unknown to them
20	CF20	Rainy or Foggy weather limiting sight. Rain and fog and disrupt the sensors ability to collect visual data. This can be hazardous even if a control action is put in place for sensors as one control action might change the whole system interaction capacity and turn into a dangerous setting.
21	CF21	Sunny (glaring light on sensors). Sensors are prone to defect when subjected to glaring light from sun
22	CF22	AV has wrong perception of road signs at the intersection approach. AV might read an old sign or does not detect a road sign.
23	CF23	AV planning and decision error. This is a component failure and should be addresed on a component reliability level
24	CF24	AV has wrong speed perception of other motorized traffic participants. This is again equipment failure and such scenario can happen no matter much of a sophisticated the system is
25	CF25	AV lacks ability to transfer non-verbal communication cues. This is related to general public not used to AVs moving around. Machines behave differently than humans and an AV moving towards a pedestrian without a human driver giving any cues of its intention to slow down or continue moving will be complicated and hazardous.
26	CF26	AV lacks ability to receive non-verbal communication cues. Similarly, Pedestrians who are in a hurry can make their way through different gestures to stop or let go an incoming vehicle. This cant pe replicated for an AV system
27	CF27	AV information system security breach. This is related to the cyber security threat with the electronic systems of AV
28	CF28	AV non-verbal communication cues ability is not effective. This is again failure of AV to effectively interpret the movement of pedestrians or even other vehicles through through human gestures which are normally used in day to day scenario
29	CF29	AV can not detect lane markings. This is failure from the sensory side
30	CF30	AV has wrong perception of cyclist and pedestrian intent. AV could not detect the pedestrian or maybe confuse it with something else.

No	Causal Factors	Description
31	CF31	AV cannot understand vehicle communication signals through use of turn signals, brake lights, hazard lights, headlights, and horn
32	CF32	AV has wrong perception of intent of other motorised traffic participants
33	CF33	AV has wrong locality perception (distance between AV and intersection). This is again component failure where sensors data and decison making component faulters
34	CF34	AV unable to transfer real-time vehicle information to AV2. A component failure or 3rd party if the communication channel is internet
35	CF35	AV is unable to give vehicle communication signals (turn signals, brake lights, hazard lights, headlights, and horn)
36	CF36	AV2 sends misleading communication of intent
37	CF37	AV2 performs illegal maneuver and does not yield when AV1 has right of way. This is important because AV2 can also cause a hazardous scenario given the fact that it is also making decisions real time and AVs with different manufacturers should have the same operation design domain to function in harmony
38	CF38	Delayed communication of intent by AV2. This is again a component failure
39	CF39	AV2 stopped in the middle of intersection due to mechanical problems.
40	CF40	AV2 sends misleading information about it's speed and intent
41	CF41	AV2 unable to transfer information to AV1
42	CF42	Av2 unable to receive information from AV1
43	CF43	The conventional car stopped in the middle of intersection due to a mechanical problem
44	CF44	Conventional vehicle performs illegal manouever and does not yield when AV1 has right of way. Drunk driving or in some cases lack of focus from the drive can be hazardous
45	CF45	Conventional vehicle driver does not follow traffic laws (does not yield when not having right of way)

No	Causal Factors	Description
46	CF46	Conventional vehicle user does not give feed back on traffic laws. This is important beacuse if the customer is not happy with the traffic rules, it means there is lack of respect and no motivation to follow them
47	CF47	Delayed communication of intent by conventional vehicle driver.This can happen in any situation. Human driver drives in an unpredictable manner.
48	CF48	The conventional car does not communicate intent through use of turn signals, brake lights, hazard lights, headlights, and horn
49	CF49	The cyclist does not communicate intent through non-verbal communication cues
50	CF50	Cyclist suddenly crosses the intersection. Or pedestrian crossing the intersection not giving enough time to AV to react
51	CF51	Cyclist can not understand AV non-verbal communication cues
52	CF52	Cyclist do not give feedback on traffic laws. This could be attributed to not having knowledge for cyclist to move in the intersection or having problems identifying separate lanes for cyclists
53	CF53	Delayed communication of intent by cyclist
54	CF54	Pedestrian suddenly crosses the intersection. Kids or pedestrians crossing the road has hazardous consequences.
55	CF55	Pedestrian does not communicate intent through non-verbal communication cues
56	CF56	pedestrian can not understand AV non-verbal communication cues
57	CF57	No feedback from pedestrians on traffic laws
58	CF58	Delayed communication of intent by pedestrian
59	CF59	Emergency vehicle suddenly passing through the intersection. These vehicles have the right of way and AV systems should be able to identify these emergency vehcicle so that they can act accordingly
60	CF60	Construction works at intersection. Vision blocking or any confusing object might end up in a hazardous situation for the AV

No	Causal Factors	Description
61	CF61	Vehicle in front does not yield (AV follows the speed of vehicle in front)
62	CF62	The vehicle in front does not proceed through the intersection (due to mechanical or driver problem)
63	CF63	Road authority does not concern itself with internet facilities at intersection. Real-time communication is key here.
64	CF64	Road authority do not regularly carryout vegetation control around intersection. Trees outgrowth or other objects put by people can block the vision. Here responsibility lies with the road authority to manage and regulate accordingly
65	CF65	Authority do not take into consideration autonomous vehicles when designing and construction of intersections
66	CF66	Road authority do not regulary maintain intersection lane markings
67	CF67	Road authority do not regulary replace damaged or missing road signs
68	CF68	Inadequate iso standards due to lack of frequent updates
69	CF69	lack of autonomous vehicle design capability by manufacturer
70	CF70	Actuator component failure. This is the failure of AV in the decison making component
71	CF71	Perception component failure. This is again a sensory failure as it has many sensors working at the same time
72	CF72	Controller is fetched with wrong input data

DEPARTMENT OF ARCHITECTURE AND CIVIL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY