Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

Karine Borgerud

Combined Safety and Security Analysis on an Autonomous Passenger Ferry Using CyPHASS

Master's thesis in Industrial Cybernetics
Supervisor: Mary Ann Lundteigen
June 2022

NTNU
Norwegian University of
Science and Technology

Karine Borgerud

Combined Safety and Security Analysis on an Autonomous Passenger Ferry Using CyPHASS

Master's thesis in Industrial Cybernetics
Supervisor: Mary Ann Lundteigen
June 2022

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

Preface

This master thesis is the final product of the degree in Industrial Cybernetics at the Norwegian University of Science and Technology (NTNU). The work has been carried out through the fall of 2022 with the help of my supervisor, professor Mary Ann Lundteigen, at the department of Engineering Cybernetics, which I would like to thank dearly. The topic was suggested by Lundteigen and she has provided her time and knowledge throughout this project.

Karine Borgerud — 20.06.22

Karine Borgerud — Date

Abstract

Today's society is becoming more and more advanced in technology, to the extent of trying to replace people for the benefit of autonomy. There already exist cars that can, under some constraints, drive autonomously and now this is being brought into the maritime sector. Having a ferry run autonomously can be a huge economical advantage because the need for a high-cost bridge may decrease. NTNU has a project where they wish to build an all-electric autonomous passenger ferry, called milliAmpere, to transport people across Nidelva in Trondheim. This has both economic and environmental advantages.

The making of such a system has its concerns. First of all, since an unmanned ferry is supposed to carry passengers there are several safety concerns for both the passengers on board but also the traffic around the ferry. Since autonomous systems are a network-connected technology, concerns around security, and especially cybersecurity, are also being discussed. Occurring cyberattacks can in a worst-case scenario affect the system in such a way that an accident happens. Because of this one needs a sufficient analysis of the system that covers both safety and security.

The field of safety and security co-analyses is novel. In this thesis, we are going to apply a newly developed analysis, called CyPHASS (Cyber-Physical Hazard Analysis for Safety and Security), on the autonomous passenger ferry – milliAmpere. The method contains the building of a model that includes a physical, cyber-physical, and a cyber layer. The method also includes several checklists to aid in the identification of the safety and security scenarios. Among these checklists, there are also implemented barriers to detect the fault and prevent or reduce the likelihood of the threat/hazard. The focus of this thesis is security, and it was therefore decided to make a limited version of the CyPHASS to better achieve this goal. We also look at other forms of sources in connection to cybersecurity to help see if the method can be improved.

The results obtained from the application of the limited CyPHASS scenario builder are presented in a series of figures. Here we discovered where in the system vulnerabilities could occur, and could therefore better protect against them. We also discovered a link between the CyPHASS and the MITRE ATT&CK database, which contain information about the tactics and techniques adversaries have previously used. Because adversaries are constantly developing their tactics and techniques to intrude on the system, it is proposed that the

CyPHASS scenario builder could integrate a database like the MITRE ATT&CK in the scenario builder. This is to help with updating the checklist, within the field of cybersecurity, in a way that we do not need to individually update each checklist at regular intervals.

Sammendrag

Dagens samfunn blir mer og mer avansert innen teknologi, i den grad at det prøver å erstatte mennesker til fordel for autonomi. Det finnes allerede biler som under visse begrensninger kan kjøre autonomt, og dette bringes nå inn i den maritime sektoren. Å ha en ferge som går autonomt kan være en stor økonomisk fordel fordi behovet for en kostbare broer kan minke. NTNU har et prosjekt der de ønsker å bygge en helelektrisk autonom passasjerferge, kalt milliAmpere, for å frakte folk over Nidelva i Trondheim. Dette har både økonomiske og miljømessige fordeler.

Å lage et slikt system har sine bekymringer. For det første, siden en ubemannet ferge skal frakte passasjerer, er det flere sikkerhetshensyn både for passasjerene om bord, men også for trafikken rundt fergen som må tas i betraktning. Siden autonome systemer er en nettverkstilkoblet teknologi, diskuteres det også bekymringer rundt IKT-sikkerhet, og spesielt cybersikkerhet. Et cyberangrep kan i verste fall påvirke systemet på en slik måte at en ulykke skjer. På grunn av dette trenger man en tilstrekkelig analyse av systemet som dekker alle aspekter av sikkerhet.

Feltet for co-analyser som dekker disse aspektene av sikkerhet er nokså ferskt. I denne oppgaven skal vi bruke en nyutviklet analyse, kalt CyPHASS (Cyber-Physical Hazard Analysis for Safety and Security), på den autonome passasjerfergen - milliAmpere. Metoden inneholder bygging av en modell hvor det inkluderes et fysisk, cyber-fysisk og et cyber-lag. Metoden inkluderer også flere sjekklister for å hjelpe til med å identifisere scenarier som oppstår på grunn av feil. Blant disse sjekklistene er det også implementerte barrierer for å oppdage feilen, og forhindre eller redusere sannsynligheten for trusselen/faren. Fokus i denne oppgaven er cybersikkerhet, og det ble derfor besluttet å lage en begrenset versjon av CyPHASS for å kunne lettere fokusere på dette målet. Vi ser også på andre former for kilder i forbindelse med cybersikkerhet for å se om metoden kan forbedres.

Resultatene oppnådd fra bruken av den begrensede CyPHASS-scenariobyggeren presenteres i en rekke figurer. Her oppdaget vi hvor i systemet sårbarheter kunne oppstå, og kunne derfor bedre beskyttelsen mot dem. Vi oppdaget også en kobling mellom CyPHASS og MITRE ATT&CK-databasen, som inneholder informasjon om taktikker og teknikker motpart tidligere har brukt. Fordi motpartene hele tiden utvikler nye taktikker og teknikker for å trenge inn i systemet, foreslås det at CyPHASS-scenariobyggeren kan integrere en database

som MITRE ATT&CK i scenariobyggeren. Dette for å hjelpe til med å oppdatere sjekklisten, innenfor feltet cybersikkerhet, på en måte som gjør at vi ikke trenger å oppdatere hver sjekkliste individuelt med jevne mellomrom.

Contents

List of Figures

List of Tables

Abbreviations

Abbreviation		Definition
APS	=	Autonomous Passenger Ship
ASC	=	Autonomous Ship Controller
ATT&CK	=	Adversarial Tactics, Techniques, and Common Knowledge
C2	=	Command and control
CIA	=	Confidentiality, Integrity, Availability
CL	=	Cyber Layer
CPL	=	Cyber-Physical Layer
CPS	=	Cyber-Physical System
CyPHASS	=	Cyber-Physical Hazard Analysis for Safety and Security
DSC	=	Design-specific cause
ECT	=	Emergency Control Team
GNSS	=	Global Navigation Satellite System
GTST	=	Goal Tree Success Tree
H/T	=	Hazard/Threat
IACS	=	Industrial Automation and Control Systems
ICS	=	Industrial Control System
IEC	=	International Electrotechnical Commission
IMO	=	The International Maritime Organization
IoT	=	Internet of things
IR	=	Infrared
IT	=	Information Technology
LIDAR	=	LIght Detection And Ranging
MASS	=	Maritime Autonomous Surface Ships
MRC	=	Minimum Risk Condition
NCA	=	Norwegian Coastal Administration
NTNU	=	Norwegian University of Science and Technology
OT	=	Operational Technology

PL	=	Physical Layer
PV-F	=	Process Variable and Functional deviations
RCC	=	Remote Control Centre
RSC	=	Remote Ship Controller
RTK	=	Real Time Kinematic
SFI	=	Centre for Research-based Innovation
SSS	=	Shore Sensor System
STAMP	=	System-Theoretic Accident Model and Process
STPA	=	System-Theoretic Process Analysis
STPA-Sec	=	System-Theoretic Process Analysis for Security
TIS	=	Traffic Information Service
UFoE	=	Uncontrolled Flows of Energy
UFoI	=	Uncontrolled Flows of Information
UFoI-E	=	Uncontrolled Flows of Information and Energy
VPS	=	Virtual Private Network

Chapter 1

Introduction

1.1 Background

Technology is moving towards a more advanced stage and integrating autonomy in this technology is becoming more and more appealing. Autonomous systems open up to the integration of smart services and minimising human intervention, however it can be challenging to build such complex trustworthy systems [1]. NTNU is currently working on a project for an autonomous ferry called milliAmpere which is supposed to transport passengers across the canal in Trondheim [2]. This technology has economic advantages since this could replace the need for building bridges, and the need for safety and security therefore increases.

Having something run autonomously raises concern for many people and the importance of safety and security increases. Nowadays more and more systems are being connected to the internet, and because of this connection systems could be vulnerable to cyberattacks. Earlier this year it was reported that a 19-year-old hacker and security researcher had breached the security in a third-party app and was thereby able to control some features of dozens of Tesla cars all over the world [3]. This app, installed in some Tesla's, allows the car owner to track the movement of the car, unlock the doors remotely, open windows, start keyless driving, honk, and flashlights. However, he did not manage to control steering, braking or acceleration which could be extremely dangerous. This is just one example of how someone could violate a vulnerability of a system and gather personal data or cause a dangerous situation.

Because industrial systems move towards a more network-connected structure, the systems need both a safety and security analysis. This gives rise to the term co-analyses, where both safety and security are well represented. An example of a co-analysis is the newly developed Uncontrolled Flows of Information and Energy (UFoI-E) method which includes the Cyber-Physical Harm Analysis for Safety and Security (CyPHASS) scenario builder. The CyPHASS scenario builder is a method with including checklists that helps form the

different scenarios where the aim is to identify safety and security barriers. There are also safety and security co-analyses that are based on already existing analyses. An example of this is the combination of the System-Theoretic Process Analysis (STPA) and the STPA for Security. Combining these two analyses makes the STPA-Extension which considers both safety and security vulnerabilities in a system.

Having just an analysis to find the vulnerabilities of the system may not be sufficient, especially since malicious actors always try to find new ways to gain access to systems. For that reason, it may be a good idea to have additional tools to help gain knowledge from the adversaries' perspective which again can assist in the work of mitigations and use of barriers. The database MITRE ATT&CK could be such a tool since it is built upon previously known attacks. If one could integrate something like this within a co-analysis, one might be able to keep up with the continuous development in cyber activities and thereby be able to better secure systems.

1.2 Problem description and objective

The autonomous ferry milliAmpere is a quite novel technology and is still in the development and testing phase. The UFoI-E method is a newly developed safety and security co-analysis, and we want to find out if this analysis is sufficient when using a use case like the milliAmpere. The main objective of this thesis is to perform the UFoI-E method to find scenarios on the milliAmpere and find out if the method can be improved. This can be summed up by answering these four research questions:

RQ1: What attributes are important in a safety and security co-analysis and which methods are suggested in the literature?

RQ2: How is a method like CyPHASS able to provide a systematic and complete approach to uncover security-related threats with an impact on safety and related countermeasures for an autonomous ferry like MilliAmpere?

RQ3: To what extent is the CyPHASS scenario builder complete, in the sense of identifying similar threats and countermeasures e.g., by comparing to the STPA-Extension method and in industry frameworks like MITRE ATT&CK ICS?

RQ4: How can the CyPHASS method be improved in terms of graphical models to support the analysis, improved checklists, or improved organisation of the method for practical use?

1.3 Project scope

1.3.1 Main tasks

To realise the objective of the thesis, the following tasks have been performed:

1. Familiarise with and describe MilliAmpere vessel and its operation, environmental influences, and control systems
2. Identify potential cybersecurity risks using databases like e.g., MITRE ATT&CK for ICS.
3. Familiarise with the CyPHASS model and related concepts used in this method.
4. Apply the CyPHASS method to identify potential cybersecurity-related threats, their potential ability to cause harm and how these can be detected and if possible, prevented or mitigated.
5. Discuss the results in connection to the STPA-Extension.
6. Suggest improvements to how the CyPHASS method can be used, based on the results of the analysis and the comparison.

1.3.2 Research approach

This thesis is based on a literature review on the topics of autonomous ships, especially the autonomous ferry milliAmpere, and safety and security co-analyses. Based on the gathering of literature on the milliAmpere, a review of the UFoI-E method is done. These results, including gathered literature on other safety and security co-analyses, will assist in the comparison and potential of suggested improvements to the UFoI-E method.

1.3.3 Limitations

To restrict the workload for this thesis some limitations have been set for the student to be able to complete the tasks within the limited time for this thesis.

- Reviewing a safety and security analysis is normally completed in teams consisting of experts within various fields. Even in teams, these tasks can require a large number of work hours, and since the review in this thesis is only done by one student within a set time limit, the workload is reduced.
- There already exists a paper where the two safety and security co-analysis, STPA-extension and UFoI-E method, have been compared. It is therefore not necessary to review a full comparison between the two analyses in this thesis.
- This comparison does not include suggestions on barriers to prevent or mitigate the occurrence of scenarios which are described in the CyPHASS method, and therefore it is more interesting to include this in the thesis rather than a full comparison.

1.4 Thesis structure

The remainder of this thesis is structured as follows:

Chapter 2 introduces the topic of autonomous ships where the autonomous ferry milliAmpere is presented. In addition, guidelines for autonomous vessels are presented.

Chapter 3 introduces the terms safety and security. The standard IEC 62443 concerning security in Industrial Automation and Control Systems (IACS) is presented. Other sources like the Industrial Control System (ICS) Cyber Kill Chain and MITRE ATT&CK for ICS are presented to assist in the review of the analysis. Lastly, two safety and security analyses are presented.

Chapter 4 presents the limited CyPHASS scenario builder, whose focus is cybersecurity. It also contains the results from the application of the analysis.

Chapter 5 contains a discussion of the findings from the results and purpose improvements on the CyPHASS analysis.

Chapter 6 concludes and presents suggestions for further work on this topic.

Chapter 2

Autonomous ships

There is an increasing trend around the topic of autonomous vehicles and their use in the future. Regarding autonomous ships, there seems to be more and more acceptance in the marine community. This acceptance is primarily for economic reasons, where cost reduction is important. There already exist autonomous or remotely controlled platforms used at sea, where the task primarily is carriers of various measuring devices. These operations are still mostly carried out nearshore and in controlled test areas or outside of shipping routes [4].

Through the years there has been a gradual development from fully manned ships to the now discussed autonomous ships. To help with the definitions of the different levels of operation modes, Table 2.1 is presented. This applies to Maritime Autonomous Surface Ships (MASS) according to the MASS UK Code of Practice [4]. Several companies are now working on full-size autonomous ships where the goal is to obtain cargo or passenger vessel capabilities. Some of these projects will be presented in the sections below.

2.1 SFI AutoShip

Centre for Research-based Innovation (SFI) AutoShip [5] is a research-based innovation centre over 8 years where the centre shall contribute to Norwegian actors taking a leading role in the development of autonomous ships. The focus will be on the integration of situational awareness, artificial intelligence, autonomous control, and digital infrastructure. There will also be developed new business models and operational concepts with cost-effective solutions for logistics and port solutions, and lastly, they will include models and methods for risk monitoring. The centre started up in late 2020 and consists of 25 partners from the Norwegian marine industry. These partners include end-users, product and service suppliers, research institutes, universities, and the government.

Table 2.1: Levels of operating modes and their definitions for MASS [4]

Level	Name	Description
0	Manned	Vessel is controlled by operators aboard.
1	Operated	Under Operated control all cognitive functionality is controlled by the human operator. The operator has direct contact with the unmanned vessel over e.g., continuous radio and/or cable. The operator makes all decisions and directs and controls all vehicle and mission functions.
2	Directed	Under Directed control, some degree of reasoning and ability to respond is implemented into the Unmanned Vessel. It may sense the environment, report its state and suggest one or several actions. It may also suggest possible actions to the operator, such as e.g., prompting the operator for information or decisions. However, the authority to make decisions is with the operator. The unmanned vessel will act only if commanded and/or permitted to do so.
3	Delegated	The unmanned vessel is now authorised to execute some functions. It may sense its environment, report its state, and define actions, and report its intention. The operator has the option to object to intentions declared by the unmanned vessel during a certain time, after which the unmanned vessel will act.
4	Monitored	The unmanned vessel will sense its environment and report its state. The unmanned vessel defines actions, decides, acts, and reports its action. The operator may monitor the events.
5	Autonomous	The unmanned vessel will sense its environment, define possible actions, and decide and act. The unmanned vessel is afforded a maximum degree of independence and self-determination within the context of the system's capabilities and limitations. Autonomous functions are invoked by the onboard systems on occasions decided by the same, without notifying any external units or operators.

2.2 AUTOSHIP

AUTOSHIP [6] is a project within the EU that aims to speed up the transition toward the next generation of autonomous ships in Europe. They will build and operate two autonomous vessels that can demonstrate their operational capabilities in Short Sea Shipping and Inland Water Ways scenarios.

This new technology with autonomous ships will help ship operators and owners in improving the economy in their investments. A reduction in ship transport can assist with the competition of replacing road transport. The aim is to increase the safety, security, and speed of the operations with the help of interoperability and Internet of Things (IoT) devices. The ships will contain advanced technology for monitoring, data fusion and communication with an evolved network.

The AUTOSHIP project will develop use cases that aim to optimise efforts and investments.

This will then assist with the development of technology for autonomous ships and allow a commercial application of this technology by the end of 2023.

2.3 The Autoferry project

Autoferry [2] is a concept of a small autonomous passenger ferry that can travel in urban areas and is a flexible and environmentally friendly alternative to manned ferries or bridges. The Autoferry project aims to develop new concepts and methods that enable the development of such ferries for urban water transport. The concept is one of the nine NTNU Digital Transformation projects [7], which have the goal to pursue research on the development and application of digital transformative technology.

Currently, there is a development in Norway where the investment in reducing emissions in shipping is increasing. The Norwegian government has a goal to have fossil-free public transportation by the year 2025 and cut the emissions from domestic shipping in half by the year 2030. This, among other things, is realised by making ferries all-electric [8]. The following means have been introduced to help with this development:

- Requirements for low or zero-emission solutions for ferries in state and county procurement of transportation services.
- Public funding in the development of all-electric powered ferries.
- Funding for the establishment of charging facilities in connection with ferry routes.

As a result, according to a mapping done by the Norwegian Coastal Administration (NCA), there is a quadrupling of charging stations for ferries in Norway from the period 2019 to 2021.

The Autoferry concept builds on these principles and develops further by making the ferry autonomous. A large part of the Norwegian population lives around coastal areas, and due to high crew costs, there is a limitation to the services available at several locations. By making the ferry autonomous, the reduction in the cost of the ferry services enables a new market in transportation. The reduction in crew cost also opens a window for other new business opportunities.

The Autoferry has one main hypothesis, in which these ferries can operate safely alongside other vessels in confined and congested environments such as urban water channels. To verify this, the project requires research methods that combine theory, simulations, experimental testing, and validation. The project consists of 19 researchers from three faculties and all three NTNU campuses [2]. It is therefore a broad competence in the project consisting of control systems, autonomous systems, sensor fusion, robotic vision, instrumentation systems, communication systems, artificial intelligence, cyber security, risk management, power systems and human factors, constituting a unique multi-disciplinary project team, which is required to solve the challenges of this project.

2.4 The milliAmpere

MilliAmpere is the name of the ferry of the Autoferry project, and there are currently two versions of the ferry, as shown in Figure 2.1. The project with the associated ferries was presented by Egil Eide at the maritime safety conference in 2018 [9].

(a) MilliAmpere (b) MilliAmpere2

Figure 2.1: The two full-scale versions of MilliAmpere [2]

The first (Figure 2.1a) is a prototype ferry with a scale of 1:2 (5 m long) which was christened on June 18th, 2018 [2]. This ferry was developed for concept testing and studying the behaviour of the other boat traffic. The second, shown in Figure 2.1b, is a full-scale ferry that can fit 12 people and is equipped with both sensor and communication systems.

In the following section, an overview of the ferry will be presented. At this stage, the paper will no longer separate the two versions of milliAmpere, and the ferry will from this point on be mentioned as an autonomous passenger ferry, or an autonomous passenger ship (APS). The work presented comes from several published papers in contribution to the Autoferry project where the milliAmpere has been analysed as a use case.

2.4.1 Overview of the APS

The system of the APS consists of several elements. The APS needs to operate within the desired area autonomously and connect to the docking station. In addition, there is a remote control centre (RCC) to which all information from the APS is communicated to. The RCC is also able to take control of the APS to remotely control it with the help of a remote ship controller (RSC).

The main goal for the APS is to transport passengers across the Trondheim city canals instead of building high-cost bridges. The operation area for the APS is shown in Figure 2.2 and the APS goes both directions across the canal with a route approximately 110m long. This is a trafficked area where mostly small boats and kayaks travel. It is therefore expected that the APS can communicate with, and safely navigate through traffic. The RCC

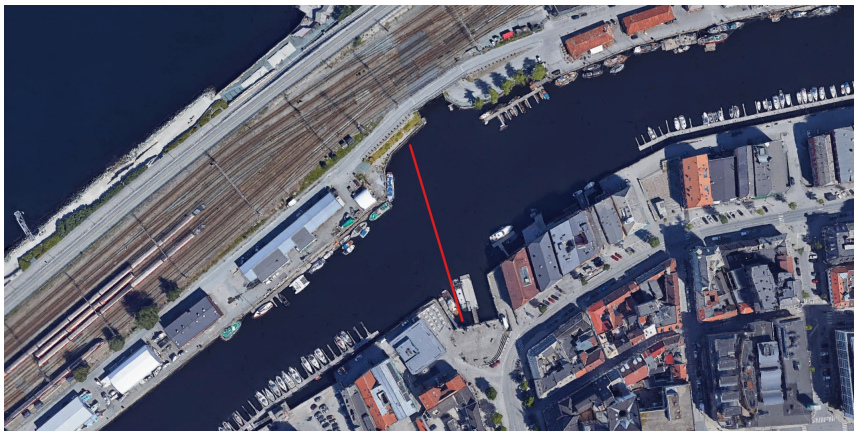


Figure 2.2: Operation area for the APS (Photo from Google Earth [10])

is planned to be located at the NTNU campus in Trondheim at an approximate distance of 1.9 km from the APS's operation area [11]. The RCC will be able to control the APS as well as monitor its operations.

To define the architecture's function and structure, Amro et al. [11] presented the system of the APS through a Goal Tree Success Three (GTST) which consists of both a Goal Three (GT) and a Success Tree (ST). The GT can be divided into three levels: the goal or functional objectives, functions, and sub-functions. The ST is a collection of subsystems to realise the functions in the GT. The resulting GTST is shown in Figure 2.3.

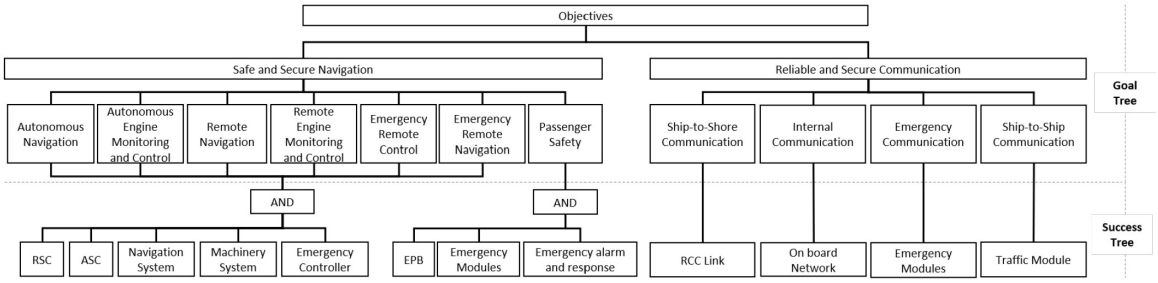


Figure 2.3: GTST of the communication architecture for the APS [11]

From the GTST in Figure 2.3 we can notice that there are two goal functions relevant to the APS, (1) safe and secure navigation, and (2) reliable and secure communication. In addition, we see the belonging functions needed to realise the goal functions.

To illustrate how the system of both the APS and the RCC connects, Amro et al. [11] have presented a context diagram which is shown in Figure 2.4. The context diagram shows that to operate, the APS depends on data from several sources, where most of these are transferred through wireless connections. These include aids to navigation, other ships, traffic information service (TIS), emergency control team (ECT), shore sensor system (SSS),

and the RCC. In this thesis, it was chosen not to include the docking system, and therefore it will not be mentioned in any further review.

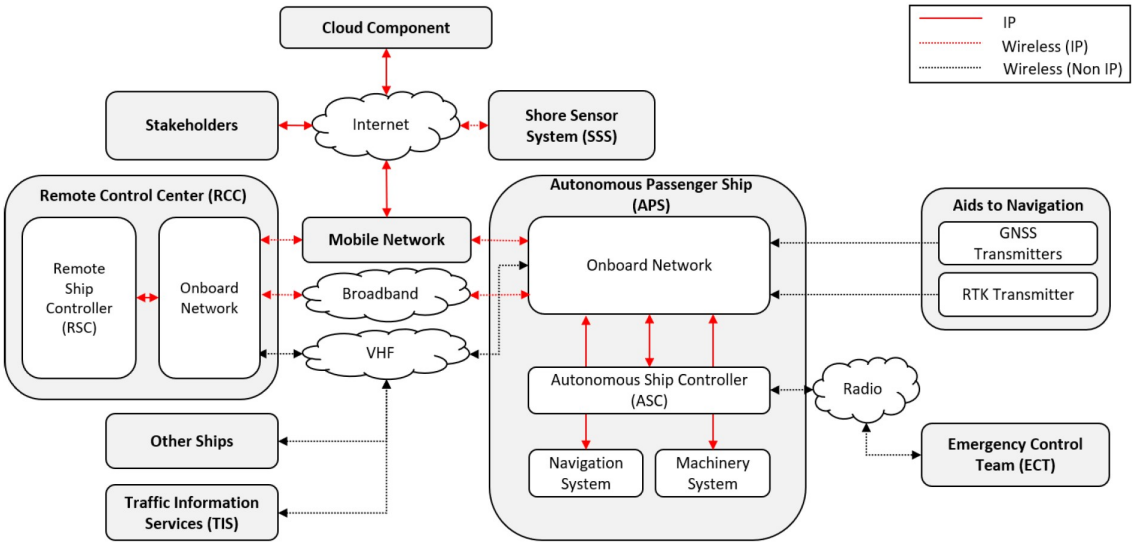


Figure 2.4: Context diagram of the APS [11]

As Eide presented in 2018 the APS will be equipped with a variety of sensors on board to aid with both navigation and safety. Here are some of the most crucial sensors on board the APS:

- **GNSS & RTK transmitter**
  The Global Navigation Satellite System (GNSS) with its internal measurement unit (IMU) have a crucial role in the navigation of the APS to provide accurate position and timing data. In addition, these components rely on the Real-Time Kinematic (RTK) receiver which provides position correction data [11].
- **Radar**
  The sensor system uses both radar and LIDAR (Light Detection and Ranging) which is a remote sensing method that uses light pulses to examine the surrounding surface [12].
- **Camera**
  Cameras aboard the APS include optical cameras with a 360 degrees range and infrared-cameras (IR).

2.5 Class guidelines

Making something autonomous, especially ships, comes with many challenges. DNV has therefore published a class guideline containing methods, technical requirements, princi-

ples, and acceptance criteria related to autonomous and remotely operated ships [13]. Autonomous and remotely controlled vessels are an area that is developing fast and there are expectations that this novel technology is implemented without severely affecting the safety of people, properties, and the environment. The International Maritime Organisation (IMO) is a specialised agency of the United Nations which has the responsibility for measures to improve the safety and security of international shipping and to prevent pollution from ships [14]. According to DNVs class guidelines, IMO does not provide any regulation for novel technologies and operational concepts such as autonomous and remotely operated vessels. However, national, and regional regulatory bodies can support the introduction of these novel technologies and operation concepts within their territorial waters. The field of autonomous and remotely operated ships is still a novel technology with an introduction of new ideas and technical solutions. Therefore, the class guidelines have been developed to support actors in the industry and the regulatory bodies in documenting and assuring a safe implementation.

DNV has an approach for assessment of autonomous and remotely operated vessels, referred to as new vessel concepts, and the following main principles from the class guideline form this foundation:

- **Equivalent safety**
  New vessel concepts shall have a level of safety that is equivalent to or better than conventional operations of vessels concerning safeguarding life, property, and the environment. The risk associated with the new operational concept shall consider consequences related to the public, assets, and environment, and not only focus on consequences for the onboard crew.

- **Risk-based approach**
  Focusing on identifying and mitigating risks associated with the newly introduced operation, functionalities and systems is necessary. There shall be performed structured risk analyses on several abstraction levels by using several different methods of risk-analyses.

- **Operational focus**
  Normally when the automated function discovers something wrong it generates an alarm in which the human operator is responsible for taking appropriate actions. Remote and autonomous operations replace human operators partly or wholly by technical arrangement. The design of the system should be done so that the need for alarm and monitoring functions corresponds to the actual possibilities for manual intervention.

- **Minimum risk conditions**
  It is important to determine how the vessel and its functions should react in the relevant situations when implementing remote and autonomous functionality. In some cases, the vessel may be forced outside its normal operation, and in such an event the relevant response should be defined. This is called minimum risk conditions (MRC) and it puts the vessel in a state where it poses the least risk to life, environment, and property.

When the autonomous and remote infrastructure experiences situations in which it cannot operate normally but still are expected to handle, the vessel should enter these MRC states.

- **Functional focus**
  Design methodology should address all the specific functions of the autoremote infrastructure to achieve the right amount of safety. Some examples of key functions of the autoremote infrastructure are remote control and supervision, communication, navigation, and manoeuvring.

- **System engineering and integration**
  This complex and new technology comes with a high focus on system engineering and integration activities. The role of the system integrator should have the responsibility for the overall functional design and verification and validation of the autoremote functionality with a focus on operation and safety for the ship.

- **Design principles**
  Several principles should guide the design of autonomous or remote ships. These are maintaining a safe state, maintaining a safe operation, redundancy and alternative control, independent barriers, self-contained capabilities on board, and self-diagnostic and supervision.

- **Software engineering and testing**
  Autonomous and remote ships are dependent on the technology of software and communication networks for control, monitoring, and safety functions on board. The integration of these complex software-based systems gives rise to the need for quality assurance for the development, delivery, and modification of these systems. Some ways of doing this are to inspect and test the end-product for defects or control the software development and configuration process to prevent mistakes from happening in the first place.

- **Cyber security**
  There is an increase in communication between the ship and remote systems. This increase is bringing concern about cybersecurity to the related systems, which is why the guidelines address this concern by emphasising the securing of the systems. This means that the infrastructure of network components, servers, operator stations and other endpoints needs to be configured such that it reduces the likelihood and consequence of cyber security breaches. This applies both on the ship and the RCC.

Chapter 3

Safety and security

Safety and security are commonly used words in organisations, industries, and even among people in everyday life. We all want to live in a safe and secure world, and these two words are therefore becoming more and more important. In this chapter safety and security will be discussed. Different tools on how to obtain, especially security, will be presented. And lastly, two safety and security co-analyses will be reviewed.

3.1 Definitions

For effective communication, it is necessary to define the terms we use. There is no right or wrong definition, only the way we choose [15]. This section is therefore dedicated to defining the terms "safety" and "security" for this thesis.

3.1.1 Safety

For at least 100 years safety has been a part of engineering. The definition of safety differs among industries, and some limit the term to only include events that impact human life and injury. To define safety Leveson [15] has used a definition that started in the U.S. Defence industry after WWII:

Definition *Safety* is freedom from accidents (losses) [15].

Definition An *accident* is any undesired or unplanned event that results in a loss, as defined by the system stakeholders [15].

Losses can include many things such as loss of human life or injury, equipment or property damage, environmental pollution, mission loss (not completing the mission), and negative business impact (e.g., damage to reputation).

Definition A *hazard* is a system state or set of conditions that, together with some (worst-case) environmental conditions, will lead to a loss [15].

In safety engineering, a hazard is defined as the state of the system, and not the environment. When eliminating losses in safety engineering one needs to be aware that some of the conditions that lead to losses are not under the control of the designer or operator (outside the boundary of the designed and operating system). To sum up, a hazard is defined as a system state that the designer and operator do not want to occur and therefore try to eliminate it, and if not possible try to control it. The goal for designers and operators is to first identify the system hazards and then eliminate or control them in the design and operation of the system.

One thing to note regarding safety is that the definition does not distinguish between unintentional and intentional causes. This is something that is more discussed in security and will be reviewed in the next section.

3.1.2 Security

Security does not relate to a hazard but rather a vulnerability i.e., a weakness in the system that leaves it open to a loss.

Definition *Vulnerability* is a flaw or a weakness in a system's design, implementation, or operation and management that could be exploited to violate the system's integrity or security policies [16].

Cybersecurity is also a term relating to security and it covers a broad range of technical, organisational and governance issues that must be considered when protecting networked information systems against accidental or deliberate threats [17]. This term is increasingly used and grows in importance since more and more businesses, governments and day-to-day activities around the world have moved online.

There are three bases in computer science that attacks on the security of information systems usually are concerned with: confidentiality, integrity, and availability.

- Breaching the **confidentiality** of systems, with data exposed to unauthorised actors.
- Undermining the **integrity** of systems, and disrupting the accuracy, consistency or trustworthiness of information being processed.
- Affecting the **availability** of systems, and rendering them offline, unusable, or non-functional.

In the thesis, the term security is mainly used to describe cybersecurity, but since there are many definitions surrounding security it includes some definitions below.

Definition *Computer security* is freedom from unacceptable risk to an information processing system where the source of harm can either be malicious or accidental [18].

Definition *IT security* deals with protecting information against unauthorised disclosure, transfer, modification, or destruction, whether accidental or intentional [18].

Definition *Cybersecurity* deals with the practice of protecting systems, networks, and programs from digital attacks. These cyberattacks are usually aimed at accessing, changing, or destroying sensitive information [19].

3.2 IEC 62443

For a long time, security was only looked at as a challenge in information technology (IT) systems. However, industrial systems are dependent on operational technology (OT) which needs to be considered when discussing security. Because of this, the series of standards International Electrotechnical Commission (IEC) 62443 were developed. The primary goal for IEC 62443 is to include security in Industrial Automation and Control Systems (IACS) through a systematic approach. IEC 62443 considers the term "security" to mean the prevention of illegal or unwanted penetration, intentional or unintentional interference with the proper and intended operation or inappropriate access to confidential information in IACS [20]. This includes computers, networks, operating systems, applications, and other programmable configurable components of the system.

The IEC 62443 series includes security for both IT and OT. This integration of the two gives a flexible framework to deal with and reduce current and future security problems in IACS. The series consists of four parts and covers the following:

- General (IEC 62443-1.* – one of four parts published)
- Policies & Procedures (IEC 62443-2.* – three of four parts published)
- System (IEC 62443-3.* – all three parts published)
- Component (IEC 62443-4.* – both parts published)

A summary of some concepts from the standard IEC 62443-1-1:2009 [16] is presented to get a better understanding of the security relating to IACS. Part 1-1 of the standard includes terminology, concepts, and models.

3.2.1 General

The environment IACSs operate in is quite complex. Sharing of information between business and industrial automation systems in organisations is increasing. Interruptions in the flow of information are no longer the only consequence of a security breach since IACS equipment connects directly to a process. This connection gives rise to injuries or potential loss of life, or production, environmental damage, regulatory violation, and can compromise operational safety.

3.2.2 Concepts

To get a better understanding of the basis of IEC 62443 we need to define some concepts concerning security. Below there are mentioned some of the concepts that have importance in this thesis, however for more detailed information, one needs to read the standard [16].

Security objectives

Traditionally IT security has focused on achieving the three objectives, confidentiality, integrity, and availability, abbreviated to CIA. In IACS this priority of the objectives often changes. The top priority in these systems is primarily maintaining the availability of all system components. Integrity is often second because of the inherent risks with industrial machinery that is controlled, monitored, or otherwise affected by IACS. Lastly, confidentiality is the least important since the data is often in raw form and needs to be analysed within context to contain any value. This prioritised order changes from IACS and general IT systems which are illustrated in Figure 3.1.

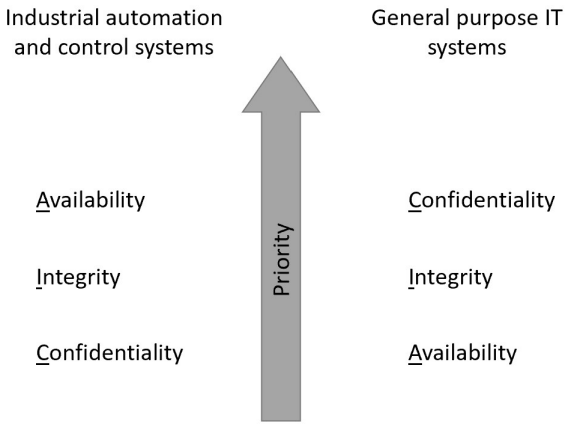


Figure 3.1: Prioritising order of objectives of IACS and general IT systems [16]

Security context

To form the basis for the interpretation of terminology and concepts, a security context is made. This shows how the various elements of security relate to each other. This shows the relationship between threats, risks, and countermeasures. These concepts and their relationship are illustrated in Figure 3.2 and show in a simple model how the interactions between the elements.

Risk

Risk is defined as an expectation of loss where risk is a function of threat, vulnerability, and consequence. Here, the consequence is the negative impact an organisation experiences

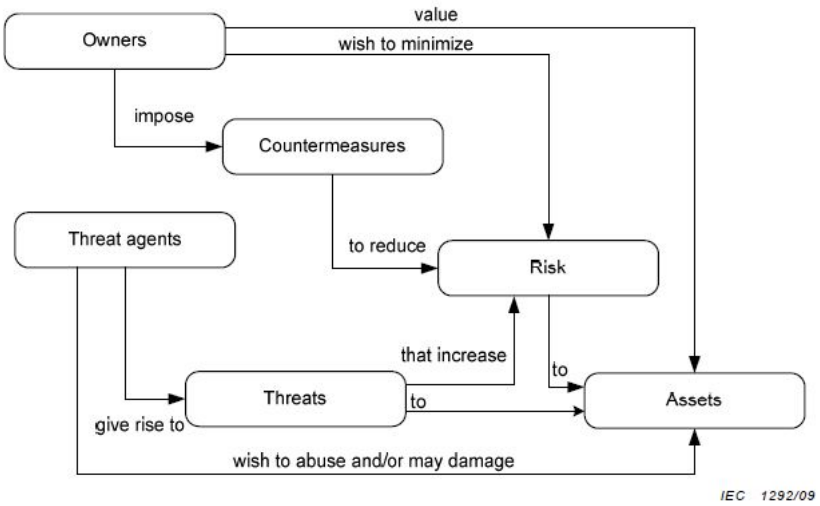


Figure 3.2: Relationship between the context elements [16]

because of the specific harm on the asset by a specific threat or vulnerability. Another way to express threat and vulnerability is in terms of likelihood (probability that a specific action will occur). The process of basic risk assessment consists of the following three steps:

1. assess initial risk
2. implement risk mitigation countermeasures
3. assess residual risk

where the second and third steps are repeated until the risk is reduced to an acceptable level. Typical risks are listed below.

- Personnel safety risks such as death or injury
- Process safety risks such as equipment damage or business interruption
- Information security risks such as cost, legal violations, or loss of brand image
- Environmental risks such as notice of violation, legal violations, or major impact
- Business continuity risks such as business interruption

Threat

Threats are described as the possible actions taken against a system. The two most common ones are accidental and non-validated changes. These threats can be defined as unintentional, however, there exist threat agents that present the threat that is looked at as intentional. These are known as **adversaries** or **attackers**, and some examples are:

- Insider: a trusted person with information that is not generally known by the public.

- Outsider: a non-trusted person or a group with inside access which may or may not be known by the organisation.
- Natural: natural events like storms, earthquakes, floods, and tornadoes are considered physical threats.

When a threat becomes an action, it is known as an attack, sometimes referred to as an intrusion. Threats can either be passive or active. A passive threat is mostly gathering information and an active threat consists of actions against and trying to harm the system/organisation.

Countermeasures

Actions taken to reduce the risk to an acceptable level are referred to as countermeasures. Typically, they do not eliminate the risk, but rather reduce it to meet the security policies. External threats can be addressed in several ways:

- Authentication of users and/or computers
- Access controls
- Intrusion detection
- Encryption
- Digital signatures
- Resource isolation or segregation
- Scanning for malicious software
- System activity monitoring
- Physical security

Internal threats need a different approach because of the attackers' possible ability to bypass some of the countermeasures such as access control. This makes it more important to include countermeasures such as written policies, separation of duties, activity monitoring, system auditing and encryption.

Threats that are quite difficult to detect are passive threats such as sniffing, which is the act of monitoring data in a communication stream. This is because this approach does not provide signals into the signal path, it only reads information moving across the connected media. Some sniffing, such as hard-connected sniffing, can be detected with help from modern communication devices, however, wireless sniffing is nearly impossible to detect. A reduction in sniffing access can be provided by controlling and closing unused voice and data ports in the plant, and by providing intelligence in communication control equipment.

Barriers

One thing that is not especially included in the IEC 62443 series, but is an important concept for this thesis, is barriers. Barriers are measures that offer protection in failure, hazard, and accident situations [21]. The barriers aim at being able to handle the risks faced at any time and are achieved by the implementation of these barriers within the relevant parts of the system. Figure 3.3 is an illustration of functions (in red) which should handle failure, hazard, and accident situations outside of the normal operation.

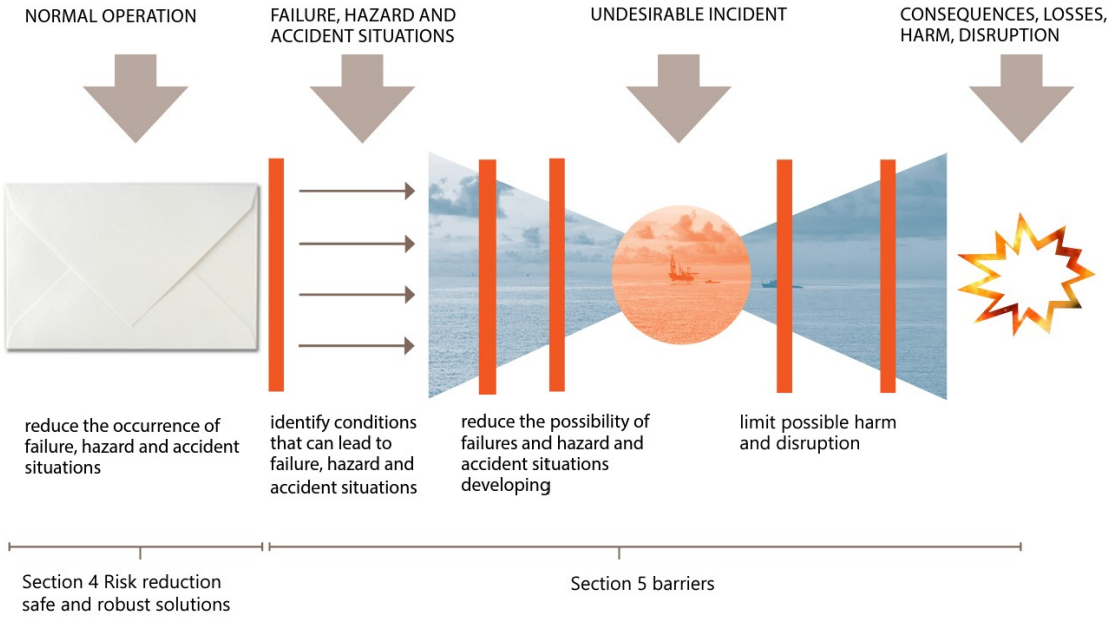


Figure 3.3: Barrier diagram [21]

3.3 Industrial Control System Cyber Kill Chain

In 2011 analysts Eric M. Hutchins et al. from Lockheed-Martin introduced the Cyber Kill Chain™[22]. This is a tool to help with the decision-making process for better detecting and responding to adversary intrusions. Based on the threat life cycle consisting of several links, breaking any of them in the chain one could prevent the attack. The concept is military-based and has been successful and used for defenders in IT and enterprise networks.

To integrate this type of concept into industrial control systems (ICS), Michael J. Assante and Robert M. Lee introduced the concept of the ICS Cyber Kill Chain [23]. A cyber attacker targets through a campaign of efforts to enable access and provide information to devise an effect. These attacks on ICS differ in impact on both the adversary's intent, their sophistication and capabilities, and their familiarisation with ICS and automated systems. Based on the understanding of where an adversary is in the campaign, it could enable the

defenders to make better decisions about security and risk management. The ICS Cyber Kill Chain is presented below with its two main stages.

Stage 1: Cyber Intrusion Preparation and Execution

This first stage deals with the adversary aiming to gain access to information about the ICS. The attacker tries to learn the system and provide mechanisms to defeat internal perimeter protections or gain access to the production environment. A summary of all the phases in stage 1 is illustrated in Figure 3.4.

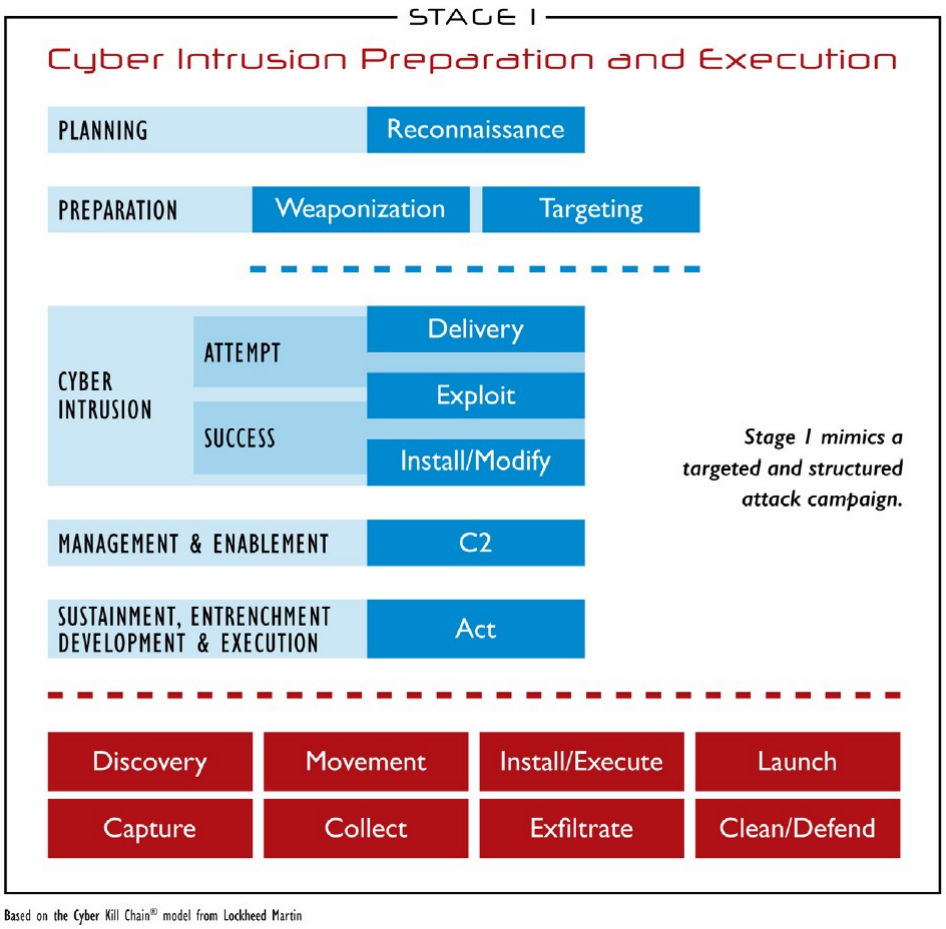


Figure 3.4: Stage 1 of the ICS Cyber Kill Chain [23]

- **Planning Phase**
  The planning phase includes reconnaissance which consists of gaining information about something through observation or other detection methods. Within cyberattacks, this phase often includes researching the target. This phase aims to reveal weaknesses and identify information that support the attackers in their effort to target, deliver and exploit elements of a system. An example of a reconnaissance technique

is footprinting, where the adversary gathers available information on the Internet and is taken advantage of it to develop information about the target without being noticed.

- **Preparation Phase**
  The preparation phase includes weaponizing or targeting. Weaponizing consists of modifying a usually harmless file to enable the next step for the adversary. This is often manifested in a file, such as a PDF, that contains an exploit. Targeting includes decisions on which attack tools or methods to use against the chosen target. These tools or methods are chosen based on trade-offs between the effort required over a period, the likelihood of success and the risk of detection. This can, for example, be discovering a virtual private network (VPN) at the defender's network that may be the best approach for the adversary and require the least number of resources. Both weaponizing and targeting can take place, but both are not required.

- **Cyber Intrusion Phase**
  This next phase, Cyber Intrusion, includes any attempt to gain access to the defender's network or system by the adversary. Both successful and unsuccessful attempts are included in this phase. The phase is divided into three steps: delivery, exploit and install/modify. The delivery step is the method by which the adversary interacts with the defender's network. This could be done with a phishing email, which is a forged email to lure the receiver to click on a website or open a document containing malicious content [16]. The exploit step is the means the adversary uses to perform malicious actions. When the exploitation becomes successful the adversary will install a capability or modify existing capabilities.

- **Management and Enablement Phase**
  When the adversary is successful in the intrusion phase, they move on to the next one, Management and Enablement, where the actor will establish command and control (C2). C2 is the set of tools and techniques that attackers use to maintain communication with compromised devices following initial exploitation [24]. Since there is a possibility that one of the C2 paths is to be detected and removed, actors often establish multiple C2 paths to ensure connectivity.

- **Sustainment, Entrenchment, Development and Execution Phase**
  This last phase documents several end goals the adversary might have. This is the phase where the adversary acts and some of the most common actions from the attacker are the discovery of new systems or data, lateral movement around the network, installation and execution of additional capabilities, launching of those capabilities, capturing transmitted communications such as user credentials, collection of desired data, exfiltration of that data out of the environment and anti-forensic techniques such as cleaning traces of the attack activity or defending his or her foothold when encountering defenders such as incident responders.

Stage 2: ICS Attack Development and Execution

In Stage 2 the attacker must now use the gained knowledge from Stage 1 and use it for intentional attacks. This knowledge must be used to develop and test a capability that can meaningfully attack the ICS.

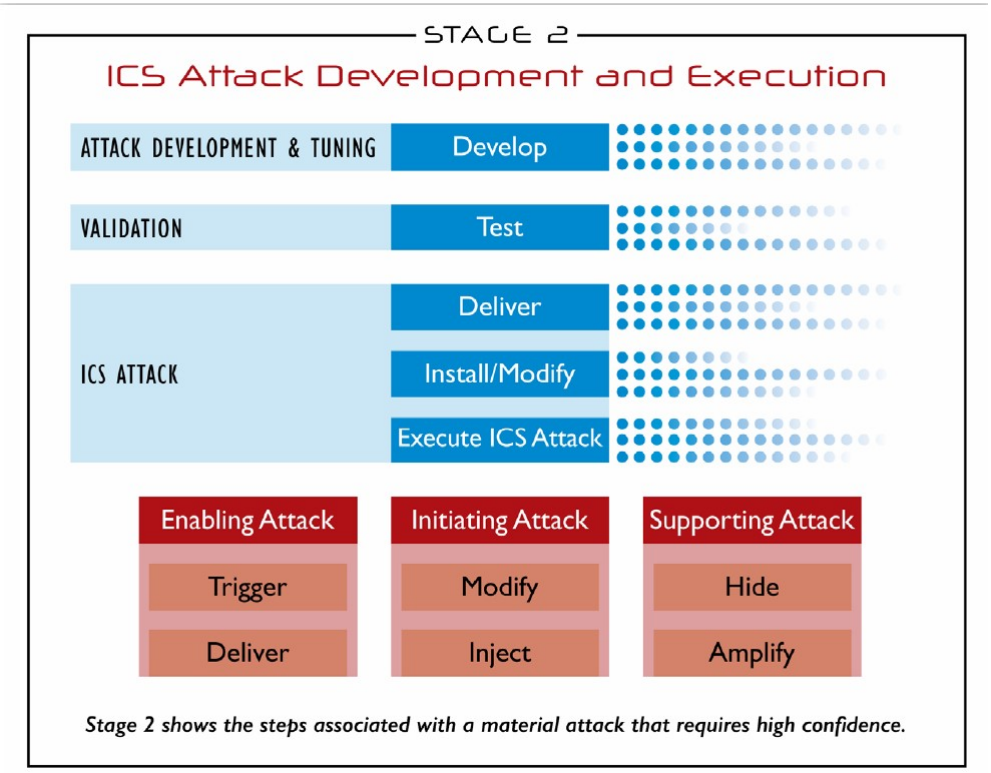


Figure 3.5: Stage 2 of the ICS Cyber Kill Chain [23]

- **Attack Development and Tuning**
  In this phase, the adversary develops a new capability to affect a specific ICS implementation. Under normal conditions, these development and tuning done by the adversary can be difficult to detect by the defenders since it is done through exfiltrated data.

- **Validation**
  After the capabilities have been developed the attacker needs to test them on an identical or similar system to check whether they have any meaningful and reliable impact. All attacks require some level of testing to confirm their impact on the ICS, even simple attacks.

- **ICS Attack**
  The last phase is the ICS Attack where the adversary delivers the capability, installs it or modifies existing system functionality, and eventually executes the attack. The

attack falls into the categories of enabling, initiating or supporting to achieve the best effect. These are used to manipulate specific elements of the process with tactics such as spoofing. Spoofing is a tactic that is often used to manipulate GNSS systems, where the receiver is tricked into believing it is at a false location [25]. There are different levels of complexity when launching an attack which is determined by the security of the system, the process that is monitored and controlled, the safety design and controls, and the intended impact. An example of a relatively simple attack is a denial of service (DoS), where authorised users are being prevented or interrupted from accessing the system [16]. This is easier to achieve than manipulation of a process and has the option to re-attack. The most common methods the adversary wants to achieve falls into the categories: of loss, denial, and manipulation. These categories include a loss of view, denial of view, manipulation of view, denial of control, loss of control, manipulation of control, activation of safety, denial of safety, manipulation of safety and manipulation of sensors and instruments.

3.4 MITRE ATT&CK for ICS

MITRE ATT&CK (Adversarial Tactics, Techniques, and Common Knowledge) is a knowledge base and model for cyber adversary behaviour [26]. It was created by the need to systematically categorise adversary behaviour, and it reflects over the various phases of the adversary's attack life cycle and the platforms they are known to target. Originally, ATT&CK was just part of a project to improve detection and malicious behaviour in Windows systems. Since then, it has grown to include other systems like Linux and macOS and has continually expanded to cover behaviour leading up to the compromise of an environment and technology-focused domains like mobile devices, cloud-based systems and industrial control systems (ICS).

Adversary behaviour in the initial stages of the attacks involving the IT infrastructure is well included in the previously mentioned ATT&CK for Enterprise. However, adversary behaviour in the later stages of these attacks is out of the ATT&CK Enterprise scope. ATT&CK for ICS was created to better understand, concentrate, and disseminate knowledge relevant to adversary behaviour in the ICS technology domain [27]. This was because that there were an increasing number of incidents associated with the attack life cycle of adversaries targeting ICS. There is a difference between the Enterprise and ICS technology domain when addressing the adversary's targets and attacks. Therefore, ATT&CK for ICS aims to include both the adversary behaviour of non-ICS technology domains and address the unique concerns of the ICS domain.

3.4.1 ATT&CK for ICS matrix

The knowledge base for ATT&CK for ICS [28] is a tool to help describe the actions an adversary may take when operating within an ICS network. There exists a matrix where the

12 tactics with their connecting techniques are visually presented. Some of these techniques can be used for different purposes and therefore connect to more than one tactic. All the tactics and its description are presented below, and the whole matrix with both the tactics and techniques can be found in Appendix A.

- **Initial access:** The adversary is trying to get into your ICS environment.
- **Execution:** The adversary is trying to run code or manipulate system functions, parameters, and data in an unauthorised way.
- **Persistence:** The adversary is trying to maintain its foothold in your ICS environment.
- **Privilege escalation:** The adversary is trying to gain higher-level permissions.
- **Evasion:** The adversary is trying to avoid security defences.
- **Discovery:** The adversary is locating information to assess and identify their targets in your environment.
- **Lateral movement:** The adversary is trying to move through your ICS environment.
- **Collection:** The adversary is trying to gather data of interest and domain knowledge on your ICS environment to inform their goal.
- **Command and control:** The adversary is trying to communicate with and control compromised systems, controllers, and platforms with access to your ICS environment.
- **Inhibit response function:** The adversary is trying to prevent your safety, protection, quality assurance, and operator intervention functions from responding to a failure, hazard, or unsafe state.
- **Impair process control:** The adversary is trying to manipulate, interrupt, or destroy your ICS systems, data, and their surrounding environment.
- **Impact:** The adversary is trying to manipulate, interrupt, or destroy your ICS systems, data, and their surrounding environment.

Within the ATT&CK for ICS knowledge base there exists a more detailed description of all the techniques adversaries may use against a system. For every technique, the knowledge base also includes suggested mitigations that can be used to prevent that technique from successfully being executed. Based on the matrix in this knowledge base it is possible to use previously gathered information on cyber intrusions and attacks to improve your system and prevent an adversary to intrude.

3.5 Safety and security co-analyses

The latest technological trends are to control systems and networks to increase efficiency. For that reason, the work on both safety and security is becoming essential since a safety-critical system is not safe if it is not secure. Elena Lisova et al. published a paper in 2019 called "Safety and Security Co-Analyses: A Systematic Literature Review" [29] where they reviewed multiple publications on the topic of combining safety and security approaches. They discovered several of the identified approaches were driven by the need in the increasing development areas such as automotive. The study showed that the safety and security co-analysis is still a developing domain.

In this paper, the focus will be on two safety and security co-analyses, CyPHASS (UFoI-E method) and STPA-Extension. Sections 3.6 and 3.7 are reviewing the two methods as well as present results from a comparative study between the two methods written by Guzman et al. [30].

3.6 UFoI-E method

In this method, there exists two concepts, Uncontrolled Flows of Information (UFoI) and Uncontrolled Flows of Energy (UFoE), where the cascade of UFoI into UFoE can be linked to the dependencies of safety and security sources of risk that could lead to physical harm. This also gives rise to the fact that the cyber, cyber-physical and physical processes are dependent on each other and interact with their environment [31]. Therefore, the concept of Uncontrolled Flows of Information and Energy (UFoI-E) is introduced, where the focus is to enhance safety risk analysis with the notion of security for safety. There are developments in both physical and cyberattacks that could lead to physical harm to such systems.

The UFoI-E method is an integrated safety and security analysis presented by Carreras Guzman et al. [32], [33], and the method consists of three constituents, 1. the UFoI-E causality concept, 2. the system representation and 3. the CyPHASS scenario builder which is illustrated in Figure 3.6. The following parts of this section will introduce the three constituents of the UFoI-E method based on the papers from Carreras Guzman et al.

3.6.1 UFoI-E causality concept

The UFoI-E causality concept integrates safety and security framework from physical, control and computer systems into a common framework. It provides the terminology of safety and security in addition to a model that abstracts the casual chain in physical harm scenarios. The view of the UFoI-E causality concept is that cybersecurity threats or software flaws can eventually cause physical damage and kill people. The control of information flows then becomes a critical requirement to prevent scenarios of physical harm.

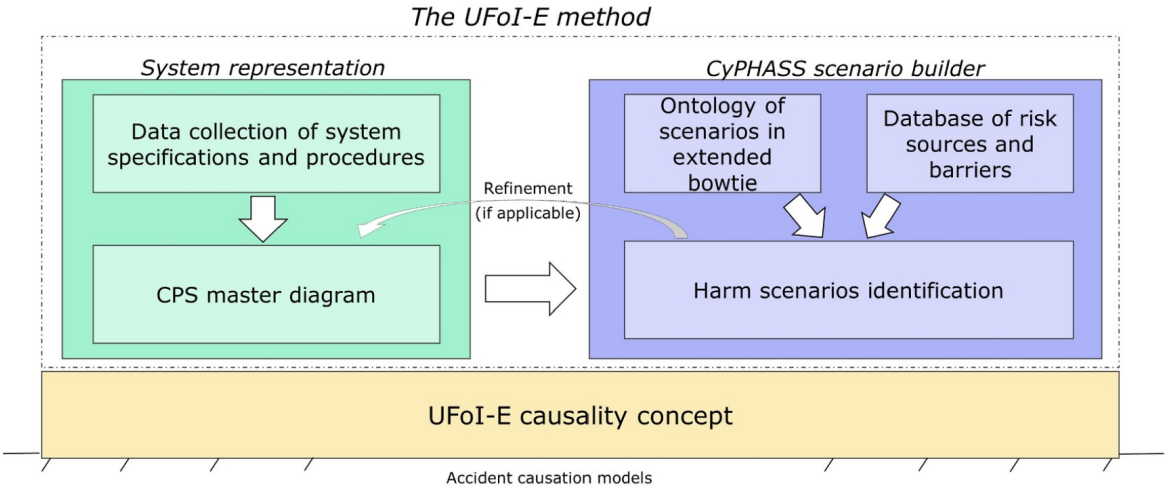


Figure 3.6: UFoI-E method as an integrated safety and security analysis [32]

3.6.2 System representation: The CPS master diagram

It may not be straightforward to analyse a complex system, and for that reason, risk analysts acknowledge that it is not the system as such that is being analysed but rather a conceptual model of the system [34]. Because of this limitation, a challenge occurs when it comes to representing the system in a comprehensible way for the analysts and incorporating all relevant features. Systems that are a combination of physical components and computer or cyber components are referred to as cyber-physical systems (CPS) [35]. These systems are high in complexity and can be difficult to model. There need to be a valid representation of the system components, interconnections and feedback control loops that identify the system and its interactions with the environment.

To provide a comprehensive representation, Carreras Guzman et al. [31] suggest an organisation of the CPS in a hierarchic structure of layers where each layer corresponds to the cyber, cyber-physical and physical processes. Including this, both the physical environment (PE) and cyber environment (CE) are represented in the model. When decomposing these subsystems and their interactions with each other give the process a greater detailed description.

The CPS master diagram is introduced, which is a diagrammatic multi-layered representation of CPSs as layers of process types. An illustration of this is shown in Figure 3.7. To bring this together, the CPS master diagram can represent a CPS as system layers and its environment with the interacting information and energy flows [33].

3.6.3 CyPHASS scenario builder

In this section the CyPHASS scenario builder is introduced which is the main contribution to the UFoI-E method, making the method an operational technique for a safety and security

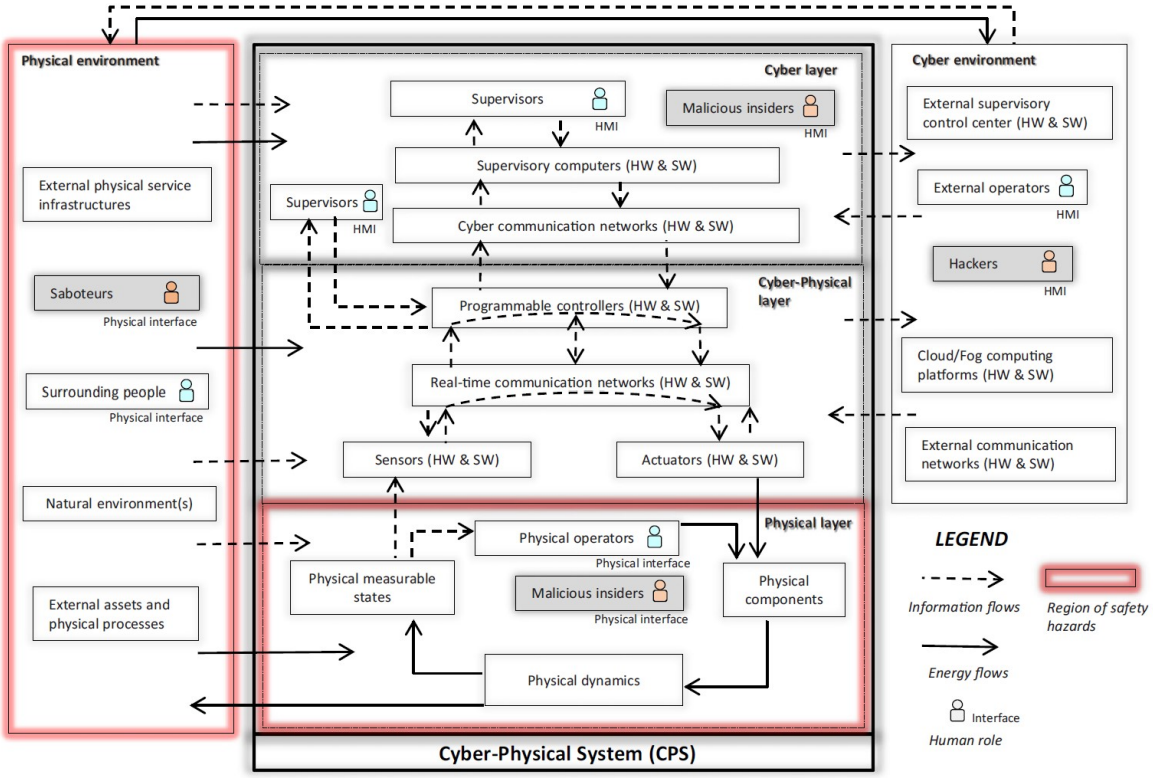


Figure 3.7: CPS master diagram representation presented by Carreras Guzman et al. [31]

analysis. Carreras Guzman et al. published a paper [32] where they present CyPHASS as a practical risk identification tool inspired by the UFoI-E causality concept and the CPS master diagram representation.

The two main components of the CyPHASS scenario builder include the ontology of scenarios visualised as an extended bowtie model and the extensive database of checklists that are embedded in CyPHASS. To assist as a toolkit, the software prototype of the CyPHASS scenario builder[1] is available as open access and is used for applying the analysis. The checklists consist of generic guidewords to help with identifying the scenarios and are built upon knowledge gathered from past events. Each block of the CyPHASS scenario builder has its checklist to provide the analyst with the help in designing the scenarios.

Firstly, the task of identifying risk scenarios in the CPS master diagram can be quite complex and therefore the ontology of scenarios provides a general framework to help with this task. Risk scenarios are a sequence of events that starts with a threat/hazard that could result in an event with potential harm. The extended bowtie model is illustrated in Figure 3.8 consisting of four top events (circles) with a set of causes to the left and a set of consequences to the right.

[1] https://doi.org/10.1016/j.ssci.2021.105458

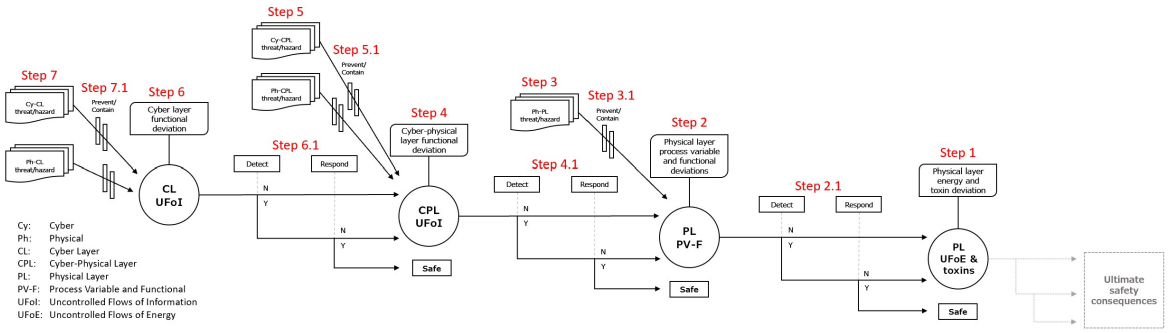


Figure 3.8: Overview of the CyPHASS scenario builder

The model is to be read backwards (from right to left) where the first top event is the UFoE at the physical layer (PL) and is a direct cause that may lead to an ultimate safety consequence. Moving on to the left of the UFoE event, the next three top events corresponds to the deviations at each layer of the CPS master diagram. The second top event corresponds to deviations at the physical layer and are possibilities of process variable and functional deviations (PV-F), the third top event to the cyber-physical layer (CPL) and the fourth to the cyber layer (CL), and these deviations are possibilities of UFoI.

Furthermore, the top events connected to the CPS master diagram are linked to an event tree, each with three branches, that describe the propagation effects between the top events. To avoid these, there are incorporated detection and response barriers. If both are present for a top event and performed according to plan, the scenario enters a safe state. However, if they are not present or get breached, the scenario from one top event propagates to the next.

Lastly, the top events are linked to some direct hazards and threats (H/T) and for that reason, the CyPHASS scenario builder presents safety and security prevention barriers to reduce the likelihood or eliminate these from happening. Therefore, if the prevention barrier is not present or gets breached, the scenario of the connecting top event occurs.

Another thing to note about the barriers in this analysis is that they are independent. The mitigation and prevention barriers act as layers of protection, meaning that if one barrier gets breached, other barriers can be activated across the layers of the system. To summarise the review of the method, the steps from Guzman et. al paper [32] are presented below:

Step 1: Identify the cases of UFoE that could lead to ultimate safety consequences in the physical layer and environment

Step 2: For each UFoE, identify the causes as PL process variable and functional deviations

Step 2.1: For each deviation in the PL, identify and recommend detection and response barriers

Step 3: For each deviation in the PL, identify causes as physical hazards and threats

Step 3.1: For each physical hazard and threat, identify and recommend prevention barriers

Step 4: For each deviation in the PL, identify causes as CPL UFoI

Step 4.1: For each CPL deviation, identify and recommend detection and response barriers

Step 5: For each CPL deviation, identify causes as cyber and physical hazards and threats

Step 5.1: For each cyber and physical hazard and threat, identify and recommend prevention barriers

Step 6: For each CPL deviation, identify causes as CL UFoI

Step 6.1: For each CL deviation, identify and recommend detection and response barriers

Step 7: For each CL deviation, identify causes as cyber and physical hazards and threats

Step 7.1: For each cyber and physical hazard and threat, identify and recommend prevention barriers

3.7 Systems-Theoretic Process Analysis

Systems-Theoretic Process Analysis (STPA), presented in the STPA-Handbook [36], is a hazard analysis technique primarily used for safety and is based on System Theory which was developed after World War II. This was to cope with the increase in complex systems with advanced technology. One of the unique aspects of System Theory is that the system is not treated as the sum of its parts, but rather as a whole. System Theory also builds on the relationship among the system parts, that is, how they interact and fit with each other.

System-Theoretic Accident Model and Processes (STAMP) [37] is an accident causation model based on systems theory and is the theoretical foundation for STPA. STAMP treats safety as a dynamic control problem rather than a prevention of failure problem. It expands from the chain of directly related failure events or component failure and includes the more complex processes with unsafe interactions between the system components. STAMP applies to any emergent properties, and because of this, STPA can be used for any system properties.

3.7.1 STPA

The STPA analysis consists of four fundamental steps presented in the STPA Handbook which are illustrated in Figure 3.9.

The first step of the analysis, *define the purpose of the analysis*, is similar to any kind of analysis method. Here one defines what kind of losses the analysis will aim to prevent, which

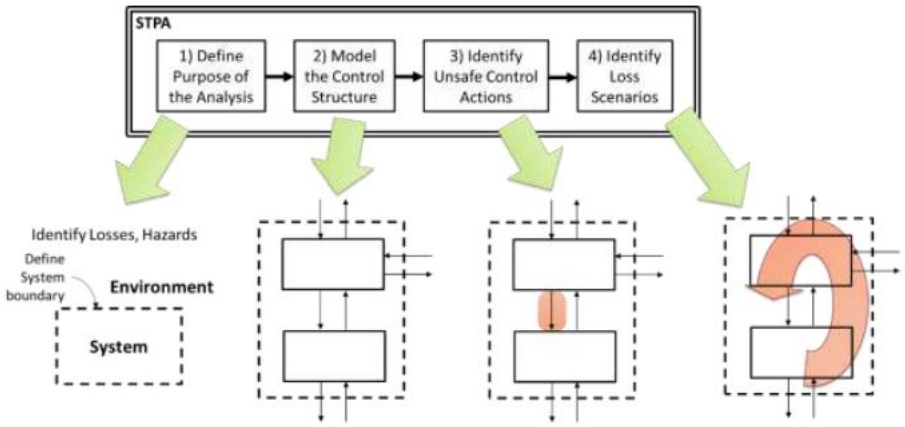


Figure 3.9: Overview of the steps in the STPA method [36]

kind of system is to be analysed, and what the system boundaries are. The second step, *model the control structure*, aims to build a model that captures the functional relationships and interactions in the system by introducing a set of feedback control loops. The third step, *identify unsafe control actions*, desire to analyse the control actions in the control structure and examine how they could lead to losses, as defined in Step 1. Based on the examination of these unsafe control actions, functional requirements and constraints are created for the system. The fourth and last step, *identify loss scenarios*, brings forth the reasons why the unsafe control might occur in the system. In the STPA-Handbook these scenarios are created to explain two things:

1. How incorrect feedback, inadequate requirements, design errors, component failures, and other factors could cause unsafe control actions and ultimately lead to losses.
2. How safe control actions might be provided but not followed or executed properly, leading to a loss.

The identified scenarios can then be used to create additional requirements, identify mitigations, and make design recommendations, among other things.

3.7.2 STPA-Sec

To include the fundamental challenges facing security, Young and Leveson have proposed a new system thinking approach called Systems-Theoretic Process and Analysis for Security (STPA-Sec) [38]. This version of the analysis is a modified version of the STPA where the aim is to address the growing problem of securing software-intensive systems against intentional disruptions. This modification of the analysis brings the focus towards a broader system structure that allows the system to enter a vulnerable system state exploited by the threat to produce the disruption leading to the loss. Like STPA, STPA-Sec shares the four basic process steps, however, the results and procedures may be different because of the focus on

security.

When identifying the unacceptable losses in the first step, these losses are most likely to extend to higher-level services rather than just physical and logical entities. Instead of using the tactic of asking *how* to best guard the network against threats, Young and Leveson's STPA-Sec uses the strategy of asking *what* essential services and functions need to be secured against disruptions. To produce a system loss, a threat needs to exploit a vulnerability. Rather than identifying all threats and then the possible vulnerabilities they might exploit, it is more reasonable to just address the system vulnerabilities that, if controlled, can prevent losses to several threats and disruptions. Control over the vulnerabilities cannot just help with the prevention of disruption of the known threats, but also the unknown threats.

Further on in the analysis, a hierarchical control structure is developed and the vulnerable control actions are identified. The potential unsafe/unsecure control actions are then identified and linked to the vulnerabilities from the first step of the analysis, and then used to develop security requirements and constraints. Finally, the last step in the analysis is to determine how the security requirements and constraints can be violated, in other words, scenarios that can lead to losses.

3.7.3 STPA-Extension as a co-analysis

It is important to be aware that safety-related causes can lead to security-related losses and security-related causes can lead to safety-related losses. Even though STPA and its modifications can analyse both safety and security related systems, they face some issues when dealing with novel cyber-physical systems (CPS). For that reason, Glomsrud and Xie have proposed a strengthening of the basic steps of the STPA as a co-analysis that is feasible for autonomous ships; a STPA-Extension [39]. They have proposed a more structured Step 1 where both safety and security are well represented. Since STPA-Sec does not explicitly consider the security-related losses, this is included in the structure to cope with that issue. In addition, a separation between security incidents and safety accidents is included to help with identifying the system-level hazards. Functional requirements are introduced in Step 1 to facilitate the development of control structures. Lastly, Step 3 & 4 have been provided with more security-related causes to further integrate security in the STPA co-analysis, whereas in Step 4 intentional causes are analysed as well as accidental and unintentional causes.

Carreas Guzman et al. have used this STPA-Extension in a study [30] on the novel concept of an unmanned, zero-emission, shortsea vessel called the ReVolt [40]. More about this study will be reviewed below. For context concerning the STPA-Extension, the study illustrated the structure of the co-analysis, which is shown in Figure 3.10.

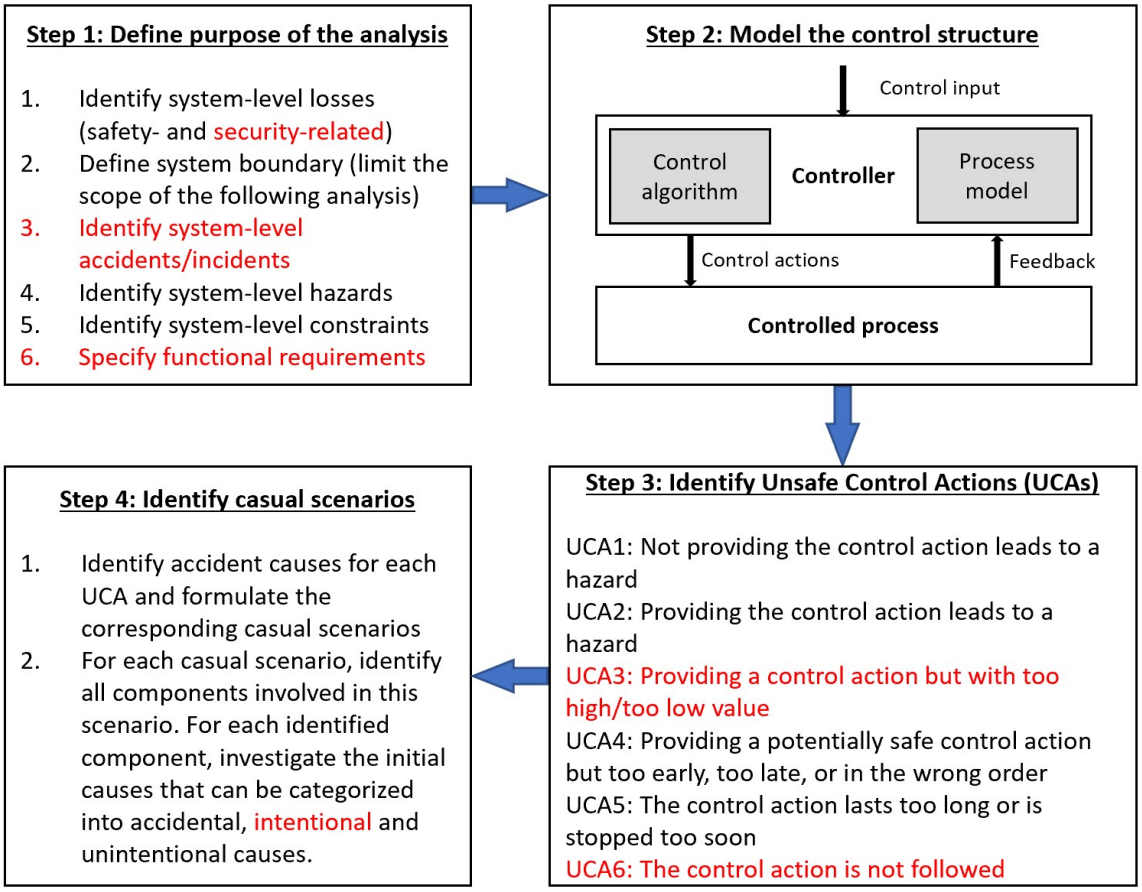


Figure 3.10: Overview of the STPA-Extension [30]. Security-related steps are marked in red

3.8 Previous work

As mentioned in the previous sections, both the UFoI-E method and the STPA-Extension have been used in a comparative study for safety and security co-analysis by Guzman et al. [30]. Both the analyses have been used on a conceptual ship with a revolutionary concept for unmanned, zero-emission, shortsea shipping called ReVolt [40]. The analyses have been done in two two-manned teams where two of the authors deployed the UFoI-E method and the other two authors used the STPA-Extension.

One thing to note in the results of the study is that the two methods have some differences, which impacts the comparison. The UFoI-E method includes prevention, detection, and response barriers in the CyPHASS scenario builder whereas the STPA-Extension on the other hand does not contain this identification and suggestion of independent barriers to prevent or mitigate the occurrence of the scenarios. These barriers were therefore not included in the comparison. This also opened up the opportunity to include and focus more on these barriers in the analysis done in this thesis.

This section presents some of the results, without going into too much detail, found by Guzman et al. to contribute to the discussion and comparison of the co-analyses later in the thesis.

3.8.1 UFoI-E method results

Firstly, the CPS master diagram was created to model the system. In this case, the diagram emphasises the autonomous mode of the system and some parts are therefore hidden. The CPS master diagram of the ReVolt can be seen in Appendix B

The results from the CyPHASS scenario builder can be summed up in Figure 3.11 whereas Figure 3.12 shows an example of the finds. As one can notice from the results in the first figure, it is that the one chosen ultimate safety consequence (collision) has a huge number of causes of propagation effects. The last figure shows an example of how a scenario can occur and propagate through the system.

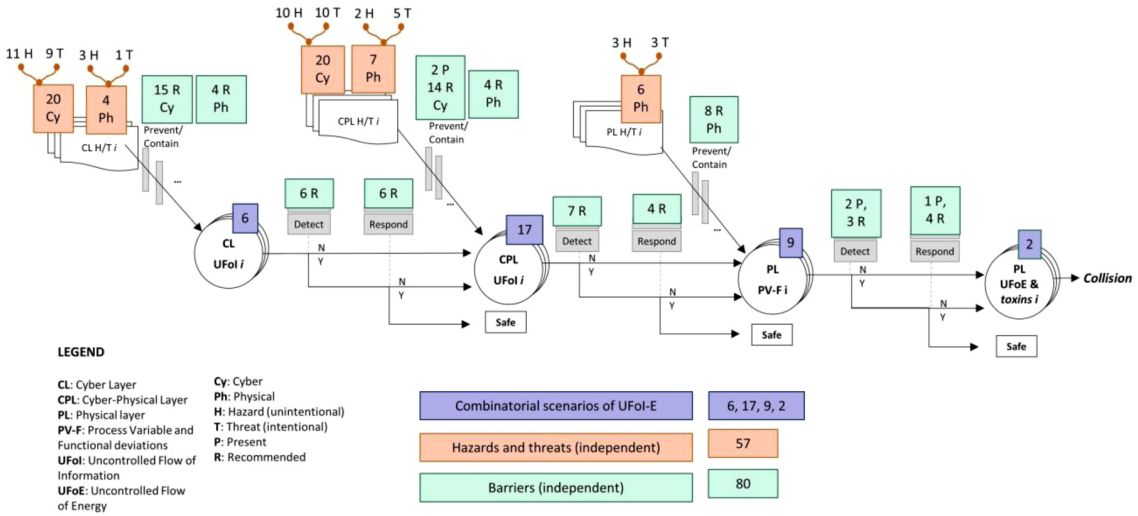


Figure 3.11: Finds of scenarios in the CyPHASS scenario builder [30]

3.8.2 STPA-Extension results

The STPA-Extension results contain a lot of information and tables, however, to not confuse the reader it was chosen to skip the first three steps and only present the results in Step 4.

By examination of the UCAs, all the potential effects of possible failures that could lead to UCAs were enumerated. This was done by following the four types of Accident Causes (AC) listed below:

- AC1 - Wrong or missing control input
- AC2 - Wrong control logic

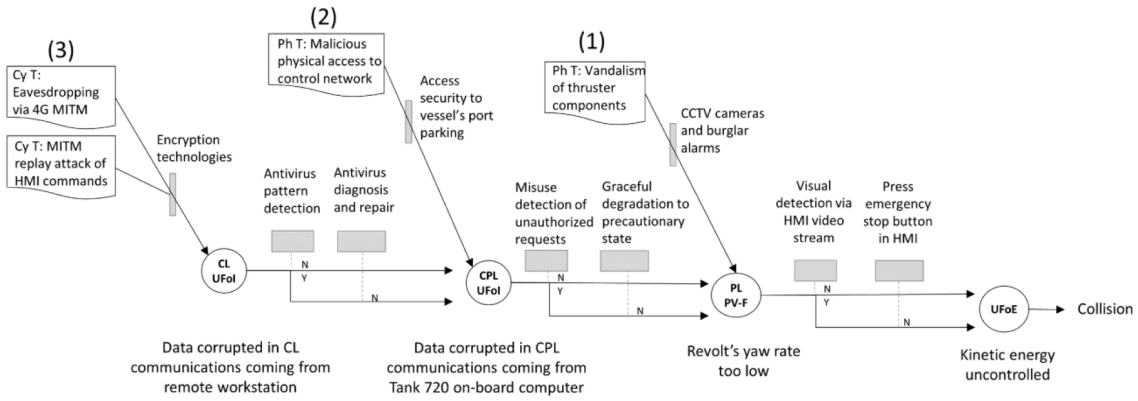


Figure 3.12: Example of scenarios with risk sources originating from different stages [30]

- AC3 - Wrong or missing situation understanding, feedback/sensing
- AC4 - Lost/altered control action or actuation

Thereby a causal scenario was specified to describe a specific effect and the resulting UCA. The given example of this is:

CS-1 (Casual scenario 1): ReVolt does not provide deviating waypoint plan (UCA1) because ReVolt does not detect or track another vessel/obstacle at all.

Table 3.1 shows the five identified Design-specific causes (DSC), which is the paper's definition of the failure modes of the casual scenario. For each of these DSCs, the initial causal factors were enumerated in the three categories: accidental, unintentional, and intentional. Here, the accidental (linked to faults in the design of an element) and unintentional (usually linked to human error) causes are in the safety category and the intentional causes are in the security category. This last cause is represented by the possible malicious entries that could exploit vulnerabilities of an element of the system. Lastly, Table 3.2 shows a summary of the causal scenarios linked to one specific hazard. Here, a single causal scenario may relate to more than one DSC, and a single DSC may be traced to more than one initial causal factor.

We have now, without too many details, looked at how the two safety and security co-analyses are being applied. Together with this and the previously mentioned theory, we will use this in applying the CyPHASS to the APS and then be able to answer the listed research questions.

Table 3.1: Design-specific causes and initial causal factors for Causal scenario - 1 [30]

DCS description	Causal factors (A: accidental, I: intentional)
Object detection model does not detect other vessel/obstacle	A1 - Object detection model fails to detect vessel/object due to its errorprone nature. A2 - On-board computer suffers hardware failure, which impacts Object Detection model. Thus, object detection model fails to detect vessel/object. I1 - Object detection model is compromised during model training phase and fails to detect vessel/object. I2 - Object detection model is attacked by adversarial inputs and fails to detect vessel/object. I3 - Object detection model is attacked on-board and fails to detect vessel/object.
LIDAR does not detect other vessel/obstacle	A1 - LIDAR suffers either hardware failure or software errors and fails to detect vessel/object. A2 - LIDAR is used in improper environment and cannot function properly. I1 - LIDAR is attacked maliciously and fails to detect vessel/object.
Obstacle avoidance module does not track detected vessel/object.	A1 - Obstacle avoidance, which has design fault, implementation errors, or suffers hardware failure and fails to track vessel/object. I1 - Obstacle avoidance algorithm is attacked maliciously and fails to track vessel/object.
Digital camera does not output picture with good enough quality or a picture at all.	A1 - RGB cameras has design fault, hardware defects or software errors. A2 - RGB camera is used in improper environment and cannot function properly. I1 - RGB camera is maliciously attacked and functions improperly.

Table 3.2: Summary of causal scenarios linked to a hazard [30]

Hazard	Number of identified causal scenarios	Number of DSC	Number of initial causal factors
ReVolt navigates too close, or at too high speed towards other vessel, object or structure	61	89	Unintentional (only related to human operator mistakes) causes: **13**, Accidental causes: **121**, Intentional causes: **104**

Chapter 4

Results

When applying the CyPHASS analysis it is interesting to see if the analysis can cover all types of scenarios. Based on the results we gain from this, we want to see if the review can uncover parts of potential improvements in the analysis. In this chapter, the results from a limited CyPHASS analysis are presented. The results are presented in a series of pictures to help with the visualisation of the analysis. Lastly, other forms of sources in connection with cybersecurity are being linked to the CyPHASS analysis to discover if it is possible to use them together.

4.1 Analysis scope: Limited CyPHASS scenario builder

The UFoI-E method is a complex method, and it can be quite challenging to complete the full analysis by oneself. To limit the workload, the steps taken in the CyPHASS scenario builder have been reduced which allows us to focus more on the cybersecurity part of the analysis. This limitation is illustrated in Figure 4.1, where the parts marked in red define the scope of our method. For our purpose, the cyber-aspect is the most important one, and it was therefore chosen to discard the physical threats and hazards to the system, as well as the uncontrolled flow of energy in the physical layer. Because of the time limit in this project and the lack of various expertise in the technical aspects of the system, the response barriers were also chosen to be discarded.

By this limitation we are left with the three main steps; the cyber layer, the cyber-physical layer, and the physical layer where the latter is mostly included to illustrate the physical scenarios that can occur. Cyber threats/hazards for both the CL and CPL are included as well as the prevention barriers for each of them. Lastly, the detection barrier is included in the scope.

This results in a reduced stepwise review of the method shown below, or in an illustrative way in Figure 4.2.

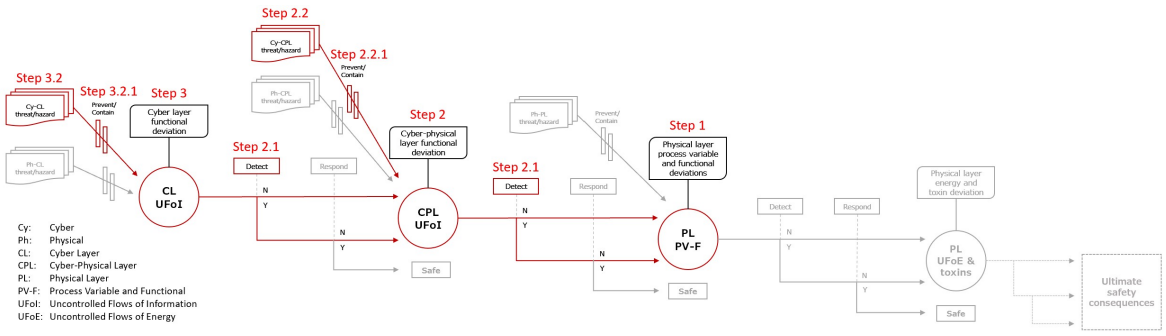


Figure 4.1: CyPHASS scenario builder with limited scope (marked in red).

Step 1: Identify the causes as PL PV-F deviations (PL)

Step 2: For each PL PV-F deviation, identify causes as CPL UFoI (CPL)

Step 2.1: For each CPL UFoI, identify and recommend *detection* barriers

Step 2.2: For each CPL UFoI, identify causes as cyber H/T

Step 2.2.1: For each *cyber* H/T, identify and recommend prevention barriers

Step 3: For each CPL UFoI, identify causes as CL UFoI (CL)

Step 3.1: For each CL UFoI, identify and recommend *detection* barriers

Step 3.2: For each CL UFoI, identify causes as cyber H/T

Step 3.2.1: For each *cyber* H/T, identify and recommend prevention barriers

4.2 CPS master diagram

Based on the literature on both the milliAmpere and autonomous ships, a CPS master diagram has been made and is illustrated in Figure 4.3. Here, the connection between the physical, cyber-physical and the cyber layer are illustrated, as well as the physical environment and cyber environment. Because of the focus on cybersecurity in this thesis, it was chosen to also include, in the CPS master diagram, areas in the process where the system could be vulnerable to cyber intrusions. These are marked with a red symbol. The methods used to attack the system are different for each stage in the process and therefore require different countermeasures to avoid the attack. Activities like phishing occur in the CL since this is where the remote supervisors operate and are receptive to these malicious emails. These remote supervisors have access to the remote ship controller (RSC) and may therefore be an appealing target for the attacker. Spoofing is concerned with signals between the cyber environment and CPL that affects the receiver in the GPS and RTK sensor to trick them into believing they are at a false position.

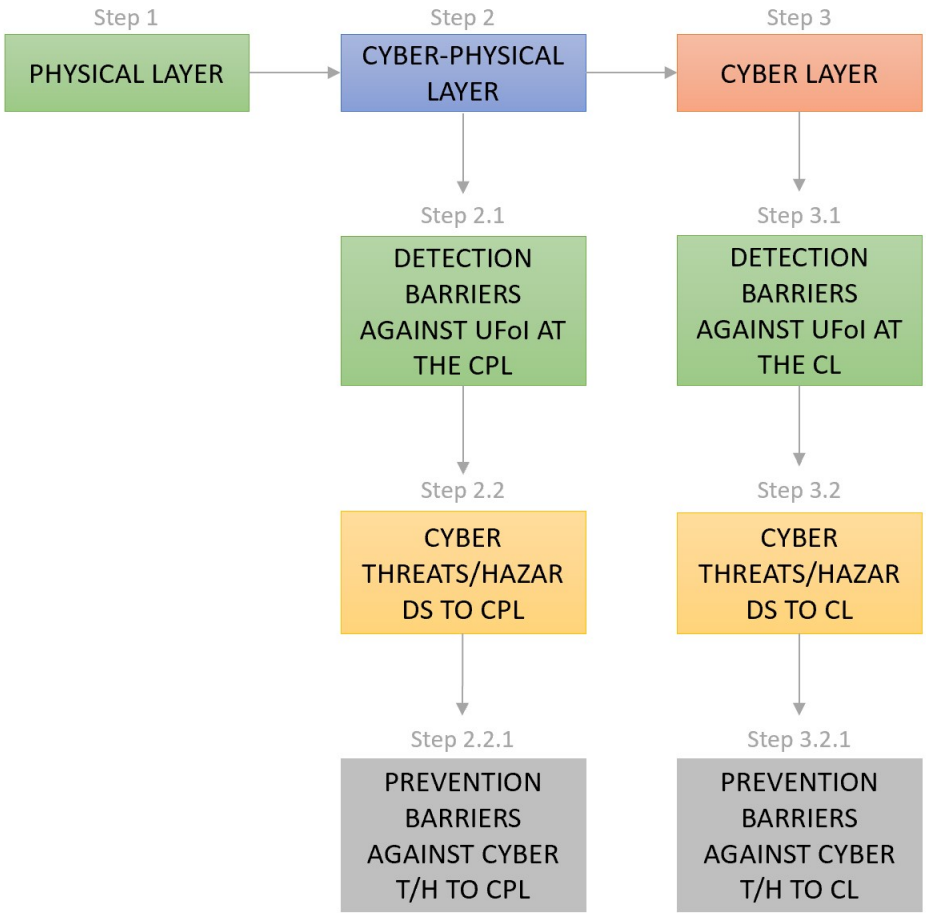


Figure 4.2: Visualisation of the steps in the limited CyPHASS scenario builder

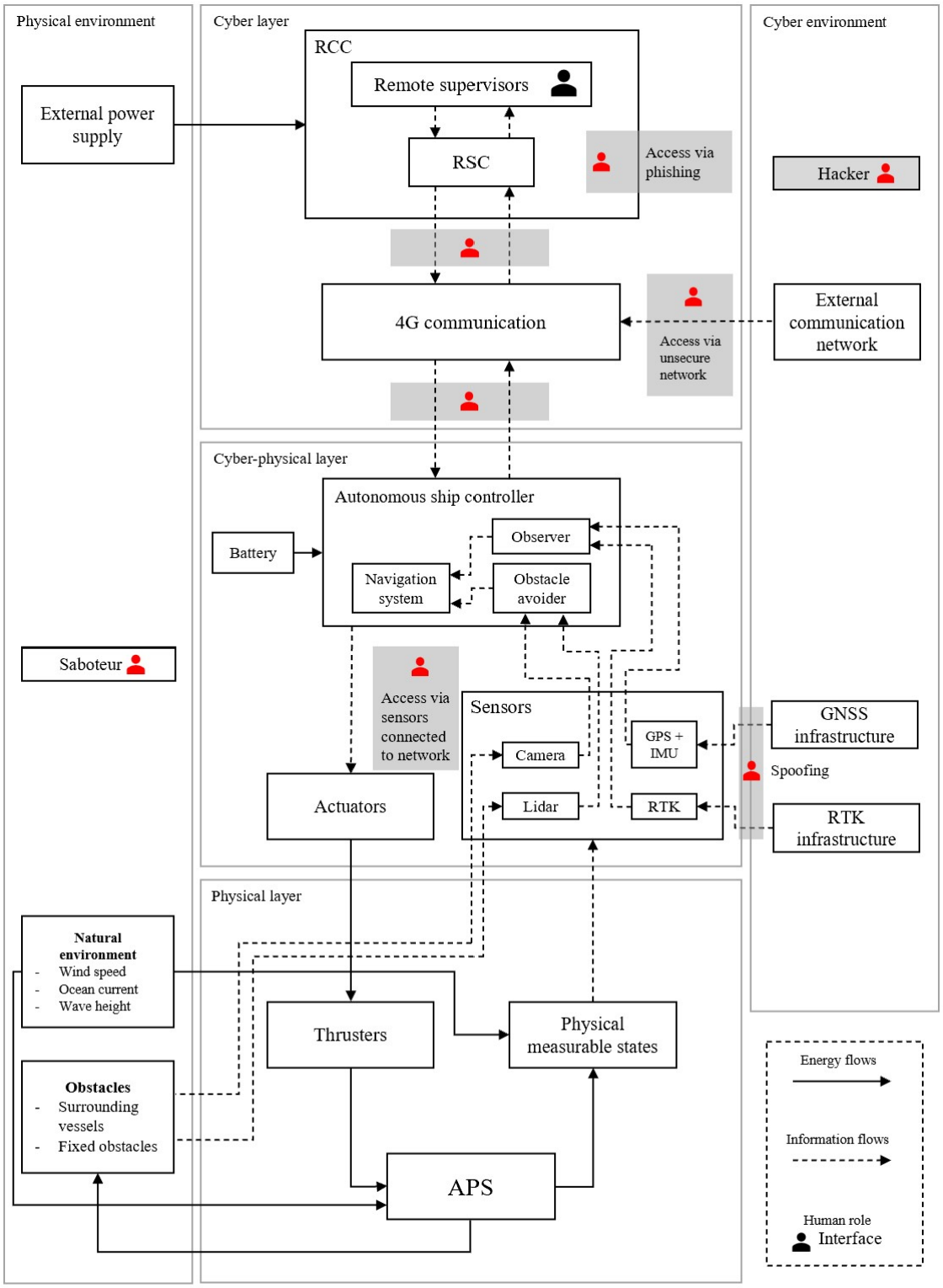


Figure 4.3: CPS master diagram of the APS

It is, in the physical environment, placed a box marked with a saboteur, which is not something that is to be mixed with the hacker in the cyber environment. The attack from a saboteur is more concerned with theft, arson, or manipulation of the external infrastructure of the physical environment of the APS. The hacker, in the cyber environment, is more interested in attacks through networks in the CL and CPL.

4.3 Scenario

For the analysis in the CyPHASS scenario builder, a single scenario as the top event in Step 1 was chosen. This was due to several factors. Firstly, this thesis's purpose is to illustrate the analysis and not do a complete analysis of the milliAmpere. Secondly, a single scenario in the PL branches out into several scenarios in both the CPL and CL. Therefore, to get a clear overview of all the steps in the analysis it was chosen to only include one scenario to not confuse the reader.

Based on the literature on autonomous boats and specific goals of the milliAmpere, the scenario in the PL was chosen to be:

- **MilliAmpere's position is wrong due to erroneous values in the navigation system**

This scenario can be thought of as a usual scenario that is desirable to avoid, especially since the navigation of the APS is one of the main elements. To repeat the element of the GTST in Figure 2.3 in Chapter 2 one of the goal functions is *Safe and Secure Navigation*. It is therefore relevant to discuss a scenario in violation of this goal function.

Figure 4.4 shows the first illustration of the implementation of the analysis where the three main steps are shown as an event tree. The one scenario from the physical layer then spreads out into seven scenarios in the cyber-physical layer. Lastly, each of the scenarios in the CPL spreads out to several scenarios in the cyber layer. As we can notice, there are quite many scenarios in both the CL and CPL that lead to a single event in the PL. Because of this, it was chosen not to include any more scenarios in the PL. In addition to this, the substeps of the analysis will only contain one scenario in the CPL and two scenarios in the CL. These are the boxes marked with a black outline in Figure 4.4.

Moving on to the sub-steps from Step 2, this can be summarised in Figures 4.5 and 4.6. From here on we look at the top event in the cyber-physical layer concerning data being corrupted in the autonomous ship controller (ASC). As we can see in the CPS master diagram, the ASC consists of several elements that can harm or disturb the operation of entering the wrong state. The controller is dependent on incoming sensor data where which is computed through the navigation system. Command actions sent to the actuators (thrusters) and communication between the APS and the RCC are crucial factors for maintaining safe and secure navigation. Therefore, corrupted data within the ASC could possibly impact several parts of the process and need to contain safety and security elements to prevent this.

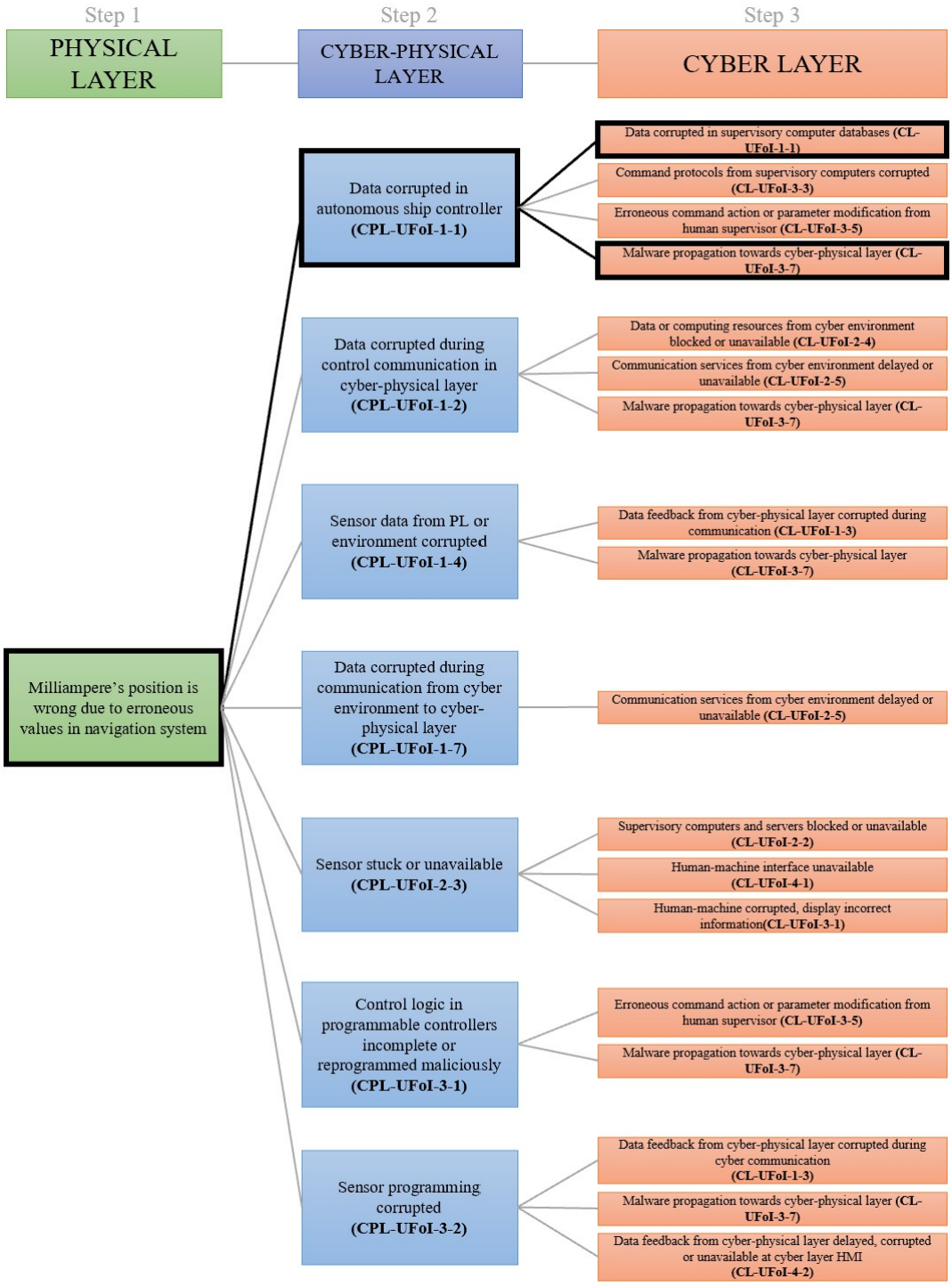


Figure 4.4: Visualisation of the three steps of the CyPHASS analysis.

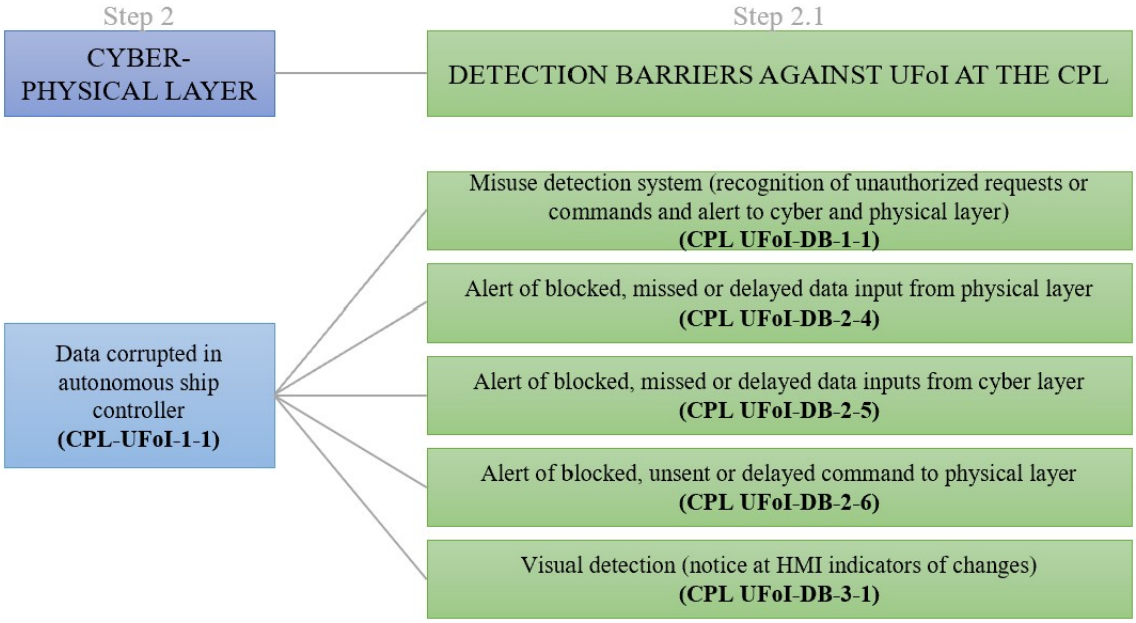


Figure 4.5: Detection barriers at the CPL

The first sub-step, Step 2.1 illustrated in Figure 4.5, is to identify and recommend detection barriers that can help with avoiding these propagation effects. A complete CyPHASS analysis will also include recommendations of response barriers where if both the detection and response barriers are present and performed according to plan, the scenario reaches a safe state. If either of these barriers is not present or gets breached, the scenario of that top event propagates on to the next.

The next sub-step, Step 2.2, is to determine the direct causes of cyber threats and hazards linked to the top event in the CPL. For this analysis, the physical threats and hazards are excluded, and we only look at cyber-related threats and hazards which are illustrated in yellow in Figure 4.6. It is important to be able to identify these threats and hazards to be able to prevent them from happening or reduce the likelihood of them. To be able to identify all of these, it is useful to use experts within the system and use previously known data of possible malicious activities.

The CyPHASS scenario builder also includes safety and security prevention barriers to eliminate or reduce the likelihood of possible direct threats and hazards. These prevention barriers have the purpose of preventing the occurrence of the related top event. Recommended prevention barriers for each of the listed cyber threats and hazards to the CPL are illustrated in grey in Figure 4.6. Here we notice that a single threat or hazard can be linked to several prevention barriers, or to think of it the other way around, a single prevention barrier can be linked to several threats or hazards. This could in practice mean that it is not necessary to implement every linked prevention barrier to cover all the direct cyber threats and hazards.

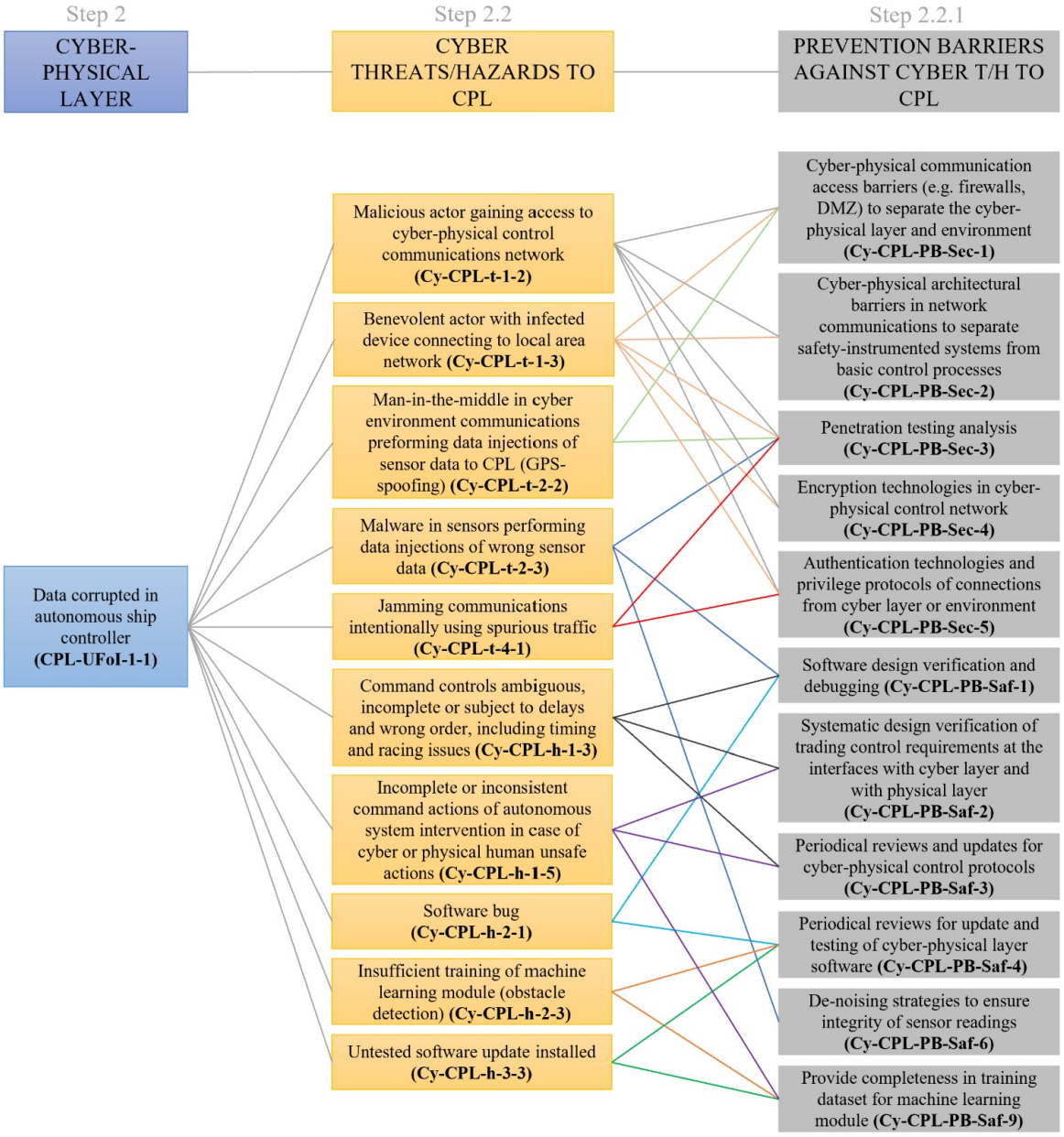


Figure 4.6: Cyber threats/hazards to CPL

Moving on to the last top event and its steps and substeps concerning the cyber layer, it was chosen to include two scenarios. The first scenario includes corrupted data in the supervisory computer databases where computer systems at the RCC could be affected. This could possibly prevent the human operators from intervening with the control of the APS when needed. The second scenario concerns malware propagation towards the CPL and could arise from several different causes. This has the intent of harming the existing system and could give rise to several safety and security problems if not detected and stopped. The reason to include two scenarios of top events in the cyber layer is to illustrate that they could connect to the same detection barriers, as illustrated in Figure 4.7. As previously discussed, similar to the prevention barriers, the system may not need all the recommended detection barriers to cover all top events and provide the system to enter a safe state.

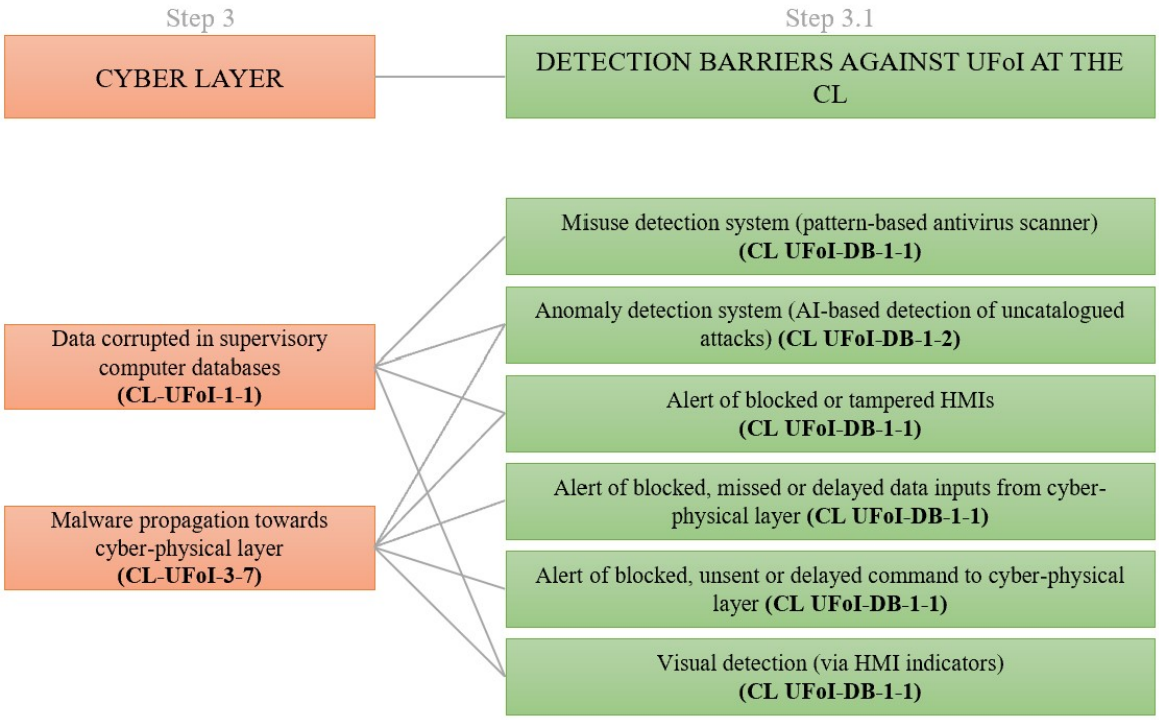


Figure 4.7: Detection barriers at the CL

Moreover, since including two CL top events we see in Figure 4.8, in the yellow boxes, they could have the same direct causes as cyber threats and hazards. Like the prevention barriers in the CPL, the prevention barriers in the CL could also be linked to several cyber threats and hazards. The chosen prevention barriers for the threats and hazards in the CL are illustrated in grey in Figure 4.8.

The CyPHASS analysis can be interpreted by its top event, where we start with the physical layer and move towards the cyber-physical and cyber layer. These represent the system deviations at each layer of the CPS master diagram. When all the top events are identified

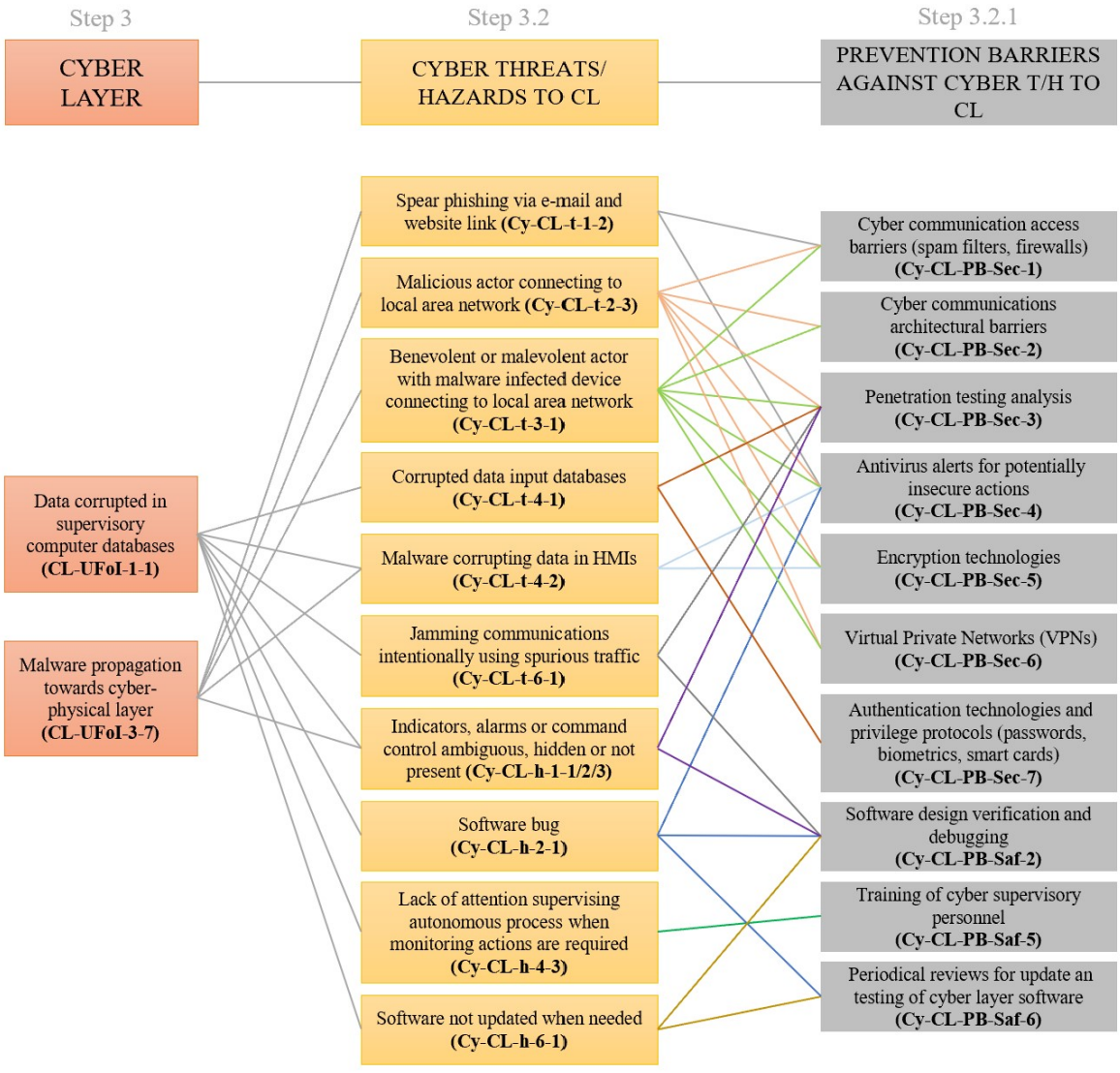


Figure 4.8: Cyber threats/hazards to CL

it is possible to read them both ways to see how they affect each others. Reading it from left to right, if the first top event occurs (red), the following top event may occur because of this. Reading it from right to left, the following top event (blue) may be the cause of why that top event (green) is occurring. Some of the chosen top events applied in this analysis are shown in Figure 4.9 to illustrate how each of them affects each other.

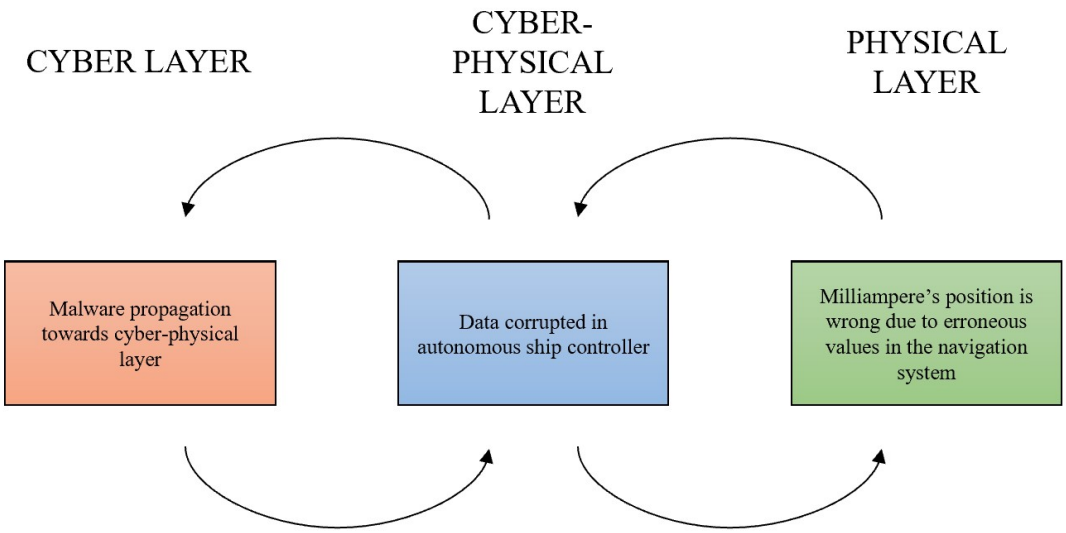


Figure 4.9: Linking between the top events

4.4 Sources of intrusion

To help with the future application of the CyPHASS analysis, it could be helpful to include other forms of sources within cyber intrusion for ICSs. It has previously been mentioned in the MITRE ATT&CK database, especially for ICS, where previously known tactics and techniques used by the adversary are presented to help prevent or mitigate them. In addition to this, the ICS Cyber Kill Chain has been mentioned where this model can assist with the decision-making of detecting and preventing attacks. Figure 4.10 is made to show a comparison between these two methods. Here we can see that each of the tactics from the ATT&CK matrix can somehow connect to steps in the Cyber Kill Chain, however, this is not the case the other way around. The three steps from the ICS Cyber Kill Chain which does not directly connect to the ATT&CK matrix include the planning phase, the preparation phase, and the validation phase. The first two refer mostly to planning prior to an actual intrusion and are therefore not explicitly mentioned as a tactic in the matrix. The last stage, the validation phase concerning testing on a duplicate system, is something that is supposedly included in most cyberattacks, however, since this part is done parallel to the system under attack it is not included in the tactics matrix.

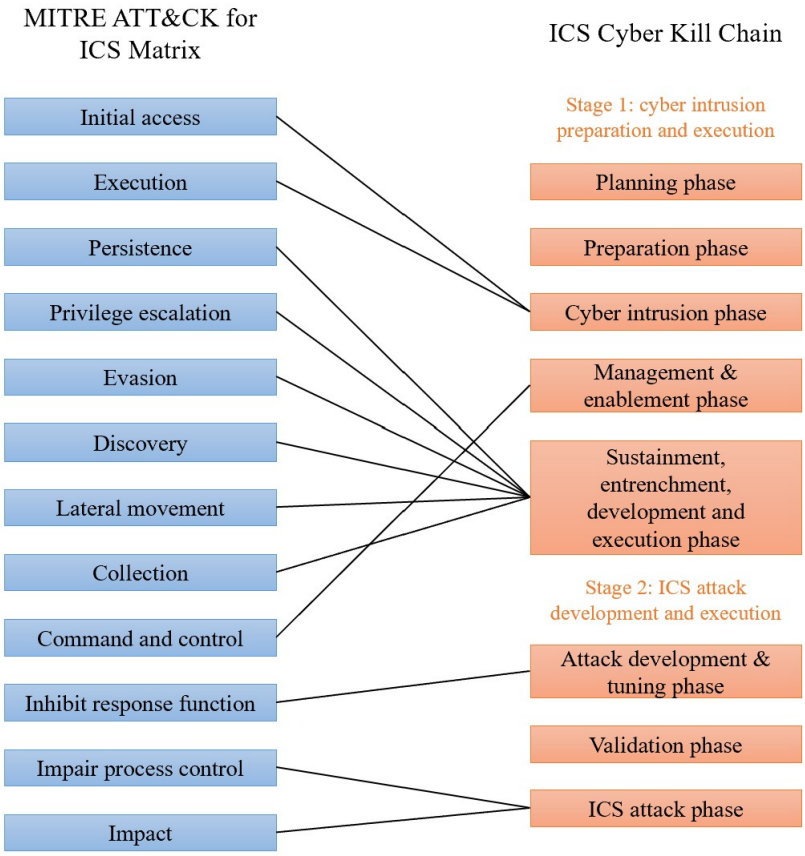


Figure 4.10: ATT&CK matrix vs ICS Cyber Kill Chain

It is also interesting to notice where in the system from the CyPHASS analysis these intrusions happen. From just the CyPHASS scenario builder, like in Figure 4.1, it could be confusing to place either part from the ATT&CK matrix or the ICS Cyber Kill Chain. Because of this, the most intuitive place to include the steps of intrusion is in the CPS master diagram which is shown in Figure 4.11. The CPS master diagram is an intuitive way of looking at the system where the physical, cyber-physical and cyber layer is included. In addition to this, the physical environment and cyber environment are included as well. As previously shown in our master diagram it is possible to illustrate more or less exactly where in the system it is vulnerable to intrusions (it depends on the level of detail in the CPS master diagram). Linking some of the tactics in the CPS master diagram corresponds quite similar to what was marked as vulnerabilities in our master diagram and using this tool in further analysis can help the analyst to discover more threats/hazards and thereby mitigate their likelihood.

Moving on from the CPS master diagram to the CyPHASS scenario builder, it is also possible to find out where one should connect the ATT&CK matrix. These tactics and techniques from the matrix are ways the adversary uses to intrude on a system. Based on this, the

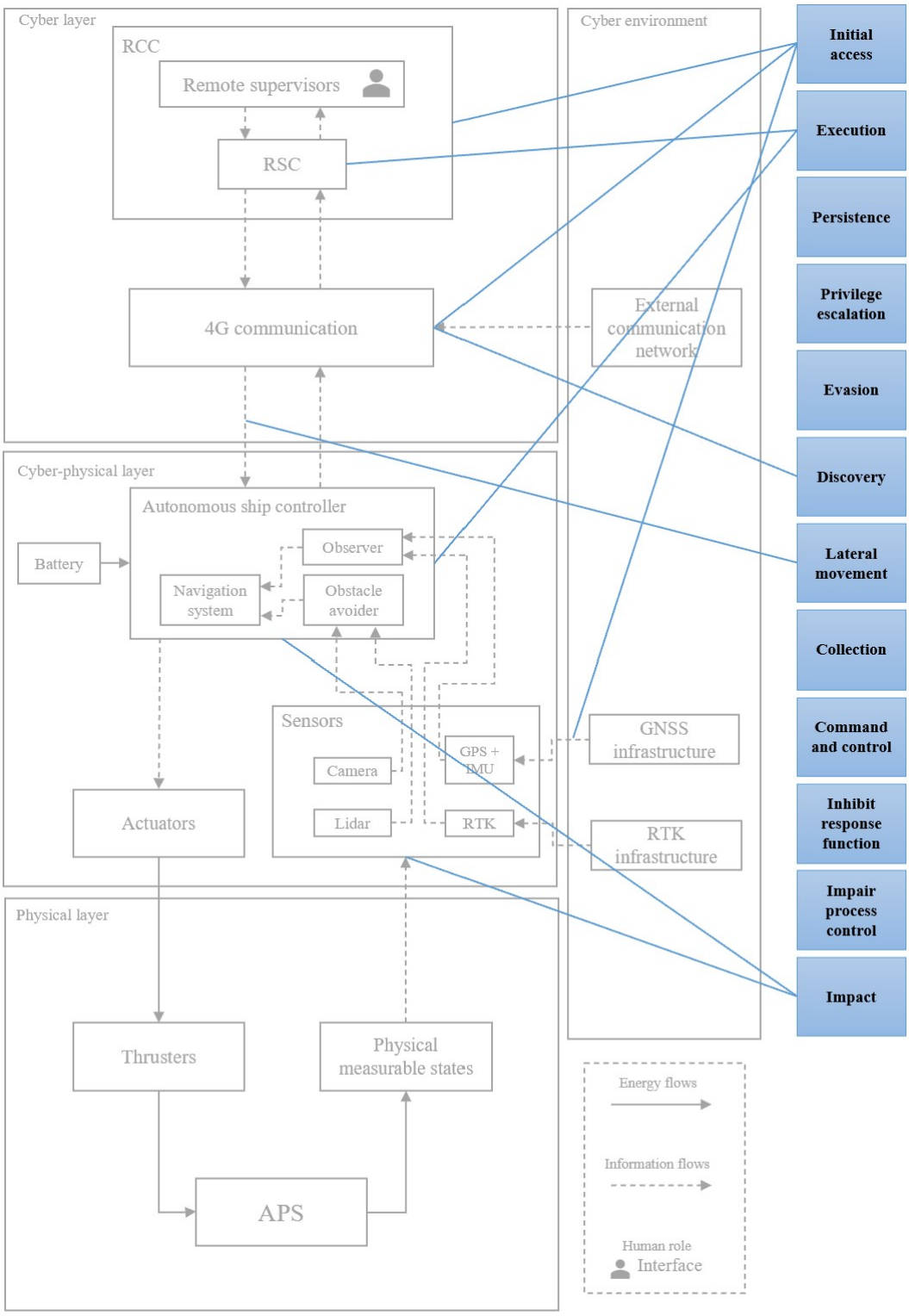


Figure 4.11: Connection between the CPS master diagram and the ATT&CK tactics

tactics and techniques should fall under the cyber threats/hazards to both the CL and CPL. The ATT&CK database also offers suggested mitigations linked to these techniques. Since placing the techniques within the cyber threats/hazards checklist, it may be a good idea to integrate these mitigations into the checklist of prevention barriers. If we were to integrate all of these in the various checklists, it is also a good idea to include a form of identification number on the techniques and mitigation to help choose the optimal prevention barrier for each threat/hazard.

To sum up, it may be useful to include databases like MITRE ATT&CK in an analysis like the UFoI-E method, both in the CPS master diagram and the CyPHASS scenario builder. This is useful because it is a way of updating the relevant checklists that is within a field that is in constant development. The challenge of this may be finding a way to integrate these things so that when new the adversary develops new methods it is easy to update both the ATT&CK database and the checklists in the scenario builder.

Chapter 5

Discussion and suggested improvements

This chapter consists of a discussion, where the results gathered from the CyPHASS scenario builder will be reviewed. It will also be given a comparison between the UFoI-E method and the STPA-Extension based on the results and review literature. In addition, we take a step back to give a summarised answer to the research questions in the thesis. Lastly, based on the knowledge gathered through literature and the experience of the review of the analysis, make some suggestions for improvements to the CyPHASS scenario builder.

5.1 CPS master diagram

It has previously been mentioned that it can be difficult to make a good model of your system with the increase in technology, especially when one is to incorporate the physical, cyber-physical and cyber layers. Through the making of the CPS master diagram, it became clearer that this model gave a better understanding of the system. It incorporated everything from the physical components to the data systems controlling them, in addition to the type of interaction between them.

The main difference to note from the hierarchical control diagram in the STPA-Extension, found in Appendix C, is that it is not as intuitive as the CPS master diagram. The STPA-Extension does not explicitly separate the physical, cyber-physical, and cyber components in its model of the control structure. However, the control diagram focuses more on how components in the system interact with each other and are therefore more detailed in that sense. One advantage of having a visual representation of your system with the interaction between the components like in the CPS master diagram is that it is possible to locate where in the system some vulnerabilities occur. This was placed in the CPS master diagram from the results in Figure 4.3. These vulnerabilities mostly occurs where components are network-connected.

5.2 Scenario builder

The scenario builder is a quite straightforward tool, and because of the help with the belonging checklists, it is manageable to do. The part that may be challenging when designing these scenarios is the lack of knowledge of the system under analysis. Therefore, with the help of experts within the fields of the system, there should be enough help to manage a complete analysis with good coverage. Doing a STPA-Extension analysis is not especially hard since it also follows a stepwise method to complete the analysis. However, because of the lack of helping material, it may be a bit more challenging to come up with the scenarios in this analysis rather than in the CyPHASS.

Also included in the CyPHASS scenario builder is the representation of barriers throughout the analysis. The detection, response, and prevention barriers play a huge role in the work that normally is done after a safety and security analysis is completed. Like in the STPA-Extension, the analysis only finds the vulnerabilities that need to be mitigated, not how they can be mitigated. It is therefore an advantage of using the CyPHASS scenario builder because the analysis is built on finding vulnerabilities in the system and right after inserting barriers to detect, respond, or prevent these. Doing such analysis on a comprehensive system can potentially be time-consuming and having the possibility to include mitigation barriers in the analysis can save time in the process after the analysis is done.

5.3 Discussion

From what is accomplished throughout this thesis the aim is now to try to answer the three first research questions, whereas the fourth and last research question is answered in the next section concerning improvements of the CyPHASS analysis.

RQ1: What attributes are important in a safety and security co-analysis and which methods are suggested in the literature?

Safety and security are terms that correlate with each other. Safety-related causes can lead to security-related losses and vice versa. It is therefore important to integrate both safety and security well when making a co-analysis. The field of safety and security co-analyses is still novel, and we have in this thesis presented two newly developed co-analyses that aim to cover both safety and security.

The STPA-Extension have the advantage of being developed from an existing analysis. The STPA-Extension is a combination of the original STPA and the STPA-Sec. The UFoI-E method is a completely new analysis, however even though the UFoI-E is a new co-analysis, based on literature, it is quite comprehensive when it comes to coverin safety and security as a whole.

RQ2: How is a method like CyPHASS able to provide a systematic and complete approach to uncover security-related threats with an impact on safety and related countermeasures

for an autonomous ferry like MilliAmpere?

Since the focus of this thesis was the UFoI-E method and especially the CyPHASS scenario builder, it was interesting to find out if the analysis could uncover security-related threats on the autonomous ferry - milliAmpere. The system of the APS is quite complex and needs to deal with a lot of safety and security-related issues.

The modelling of the system was done as a CPS master diagram. This makes a lot of sense because the model includes the three different layers which are becoming more important in newer technology. The STPA-Extension model of the control structure is not as intuitive about the system, as the CPS master diagram is, and I would therefore prefer to use the latter. The results from the CyPHASS scenario builder appeared quite comprehensive when it came to security, with the exception of details in the scenarios. If one performs the analysis in a workshop with other experts about the system, it would seem that the analysis covers security-related threats and the countermeasures needed.

RQ3: To what extent is CyPHASS complete, in the sense of identifying similar threats and countermeasures e.g., by comparing it to the STPA-Extension method and in industry frameworks like MITRE ATT&CK ICS?

From my experience of reviewing the CyPHASS scenario builder, the method appeared to be complete in the sense of doing it with little prior knowledge of the method. I was able to complete the analysis and got results that made sense. Because of the lack of expertise in this field, the results are quite generic and do not go into specific details of the system. This may be problematic when trying to discover faults in the analysis since the results are not specific enough for the analysis to work properly. Improvements to the checklists are also difficult to do because the results contain only the non-specific scenarios and barriers. With a range of more specific scenarios, it would have been easier to discover deficiencies in the analysis.

The checklists are made based on previously known events and knowledge. By using, for example, a database source like MITRE ATT&CK one could implement more knowledge about cybersecurity upon the checklists. From looking at the analysis I found that the relevant checklists in the scenario builder were cyber threats/hazards in the CL and CPL and their prevention barriers. These are similar to the techniques and mitigations lists in the database and are the most relevant to use when updating the checklists. The ATT&CK database may have the potential to be updated more frequently than the CyPHASS scenario builder about current trends of adversaries, therefore it makes more sense to be able to use the matrix within the scenario builder rather than just updating the checklists.

5.4 Improvements to the CyPHASS scenario builder

After gaining knowledge and experience about the CyPHASS analysis, and other forms of methods, it is finally time to answer the last research question concerning improvements to

the method.

RQ4: How can the CyPHASS method be improved in terms of graphical models to support the analysis, improved checklists, or improved organisation of the method for practical use?

Firstly, it is important to note that the use case was applied on a limited CyPHASS scenario builder. This limitation was very useful in this context because it allowed shifting the focus more toward cybersecurity while still being able to apply the analysis. Since showing the parts related to security, or cybersecurity, in the analysis, it also made more sense to connect the analysis to the database of MITRE ATT&CK.

One thing to notice is the way the CyPHASS analysis is constructed in its steps. The original analysis contained seven steps, plus some sub-steps, and even with the limited scope of this thesis the method still needed all these steps, except Step 1. The thesis presented a remodelling of these steps which made it only include three main steps, each of them connected to the belonging top event. This makes the steps in the method more connected to the maser diagram with its connection to the different layers, and the experience of applying the method made it clearer.

Another thing that could improve the CyPHASS scenario builder, which has previously been discussed, is to include other sources to strengthen the way one uses the checklists in building scenarios. The MITRE ATT&CK database contains information about previously known attack groups and the tactics they use. Along with these, there are suggested mitigations which can be of help when designing a system with similar issues. Inhe result chapter it was tried to integrate some of the tactics within the CPS master diagram to illustrate where in our system these could be used. Based on this, it is possible to either improve the existing checklists of CyPHASS based on the ATT&CK database or try to integrate the ATT&CK matrix to CyPHASS and use it as an extra supplementary material alongside the already known checklists.

Chapter 6

Conclusion and further work

6.1 Conclusion

Because of the growing technology, it is most likely that machines are becoming more and more autonomous, and therefore replacing manned vehicles or ships. There are several advantages of doing this and the main one is economic advantages. The autonomous ferry which is the example for this thesis has the advantage of carrying passengers across the canal in Trondheim which eliminates the cost and time of building a bridge. This issue is something quite relevant in Norway because of the infrastructure with a long coastline and many fjords. Today many people are dependent on ferries to commute to work or transport goods. As a long perspective plan, it is a huge market in the country to adopt autonomous ferries. If we also include all-electric autonomous ferries, it can reduce the environmental waste in Norwegian shipping.

Making something autonomous, especially something that deals with people, gives rise to the importance of safety and security. To integrate these two concepts, we must have a good co-analysis. In this thesis, it was used the UFoI-E method, which included the CyPHASS scenario builder as the analysis under review. In addition to this, the STPA-Extension was presented as a secondary safety and security co-analysis, and the MITRE ATT&CK for ICS was discussed on how to include this as supplementary material. With the help of the checklists for the scenario builder, it seemed easier to make scenarios for our system. The CPS master diagram also made sense when analysing such a complex system because it simplified it as a model and yet seemed to cover a lot of grounds in the system. From my experience, the CyPHASS scenario builder is a great tool to do a safety and security analysis because of all the supplementary material within the method. Because of this even non-expert like myself can apply the method.

When doing a safety and security analysis of any type of system, they are normally done in teams of several people with different types of expertise. Throughout the review of the

analysis, it became clear that the amount and variety of knowledge are crucial in completing such analysis. Even though I was able to complete the part of the analysis that was reviewed, I found it hard and time-consuming. For any further reviews of such an analysis, it is preferable to do so in a workshop. This is to gain more knowledge of every aspect of a system and gives rise to a discussion about the topic where one could uncover more safety and security concerns.

6.2 Recommendations for further work

Based on the work done in this thesis, there is still something that could be improved and worked further on. The following points are my recommendation for further work on this topic:

- Because the CyPHASS scenario builder contains a lot of information, I would suggest making a better visualisation of the results gathered from the analysis. Is there a way to organise the results in a better way?
- The MITRE ATT&CK matrix is a good tool to help with the security part of an analysis, therefore I suggest finding a way to integrate this, if possible, in the relevant checklists of the CyPHASS scenario builder.

Appendix A

The ATT&CK for ICS Matrix

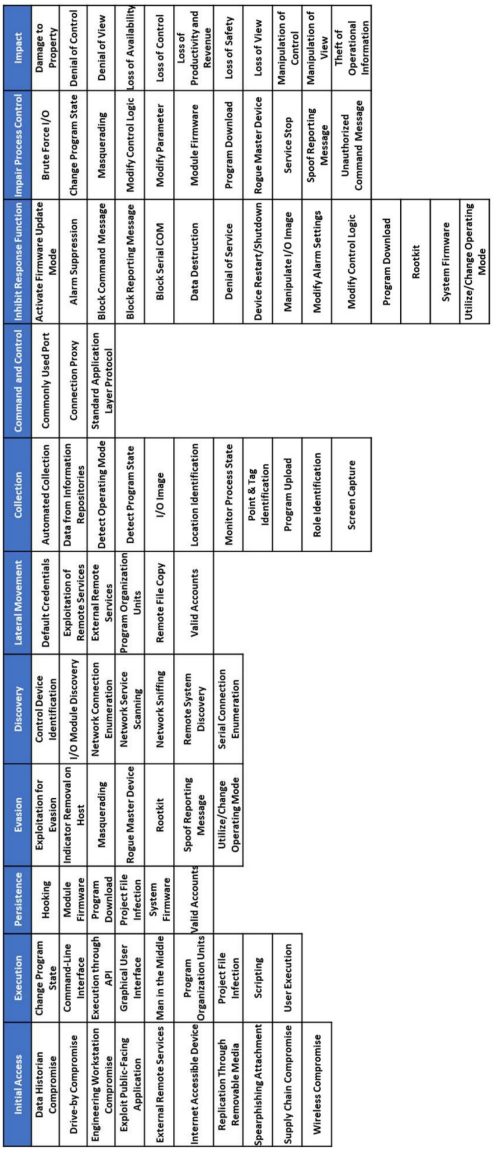


Figure A.1: The MITRE ATT&CK matrix with its 12 tactics and connecting techniques

Appendix B

CPS master diagram of the ReVolt

reference: [30]

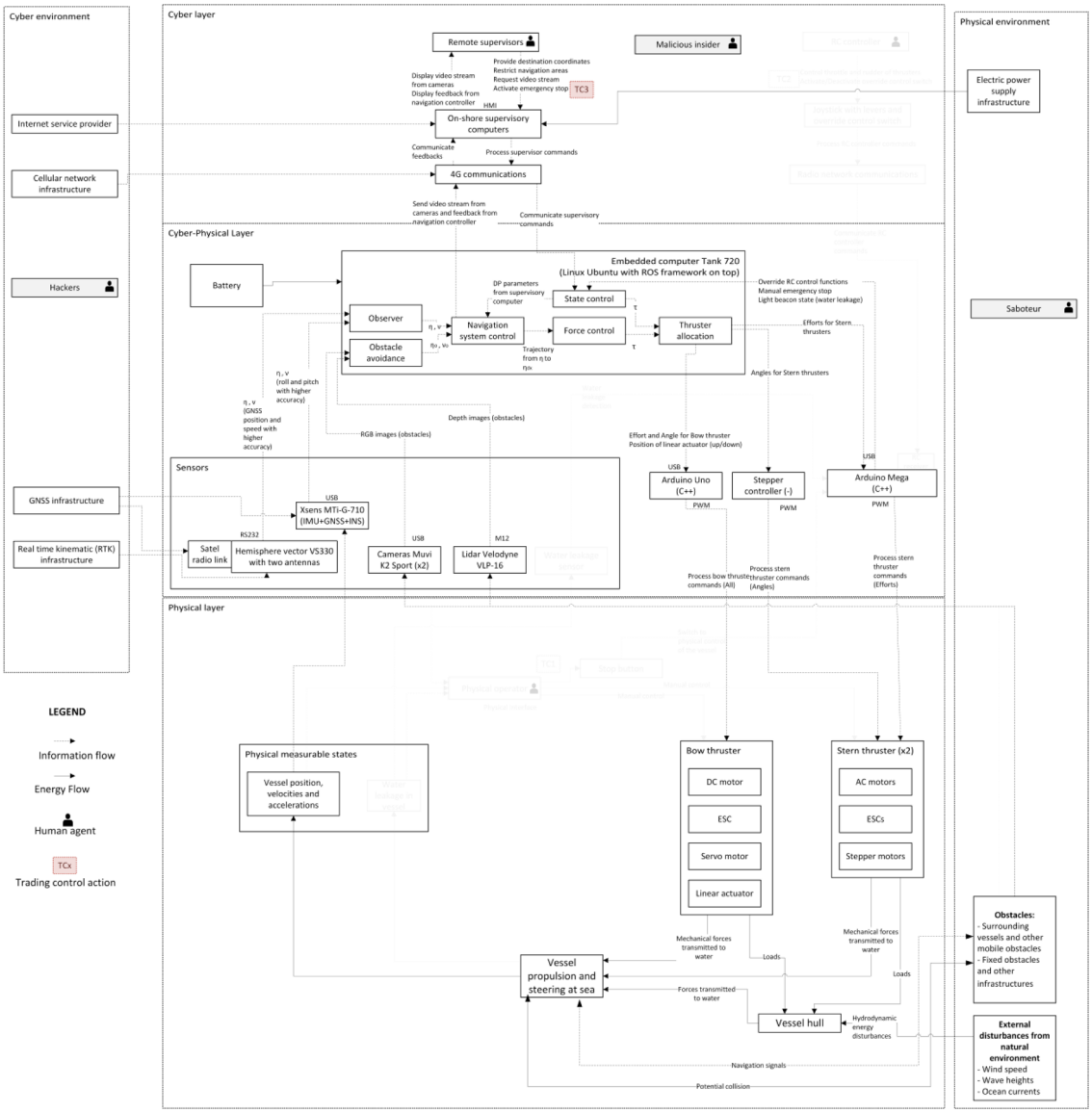


Figure B.1: CPS master diagram of the ReVolt [30]

Appendix C

Control structure of the ReVolt

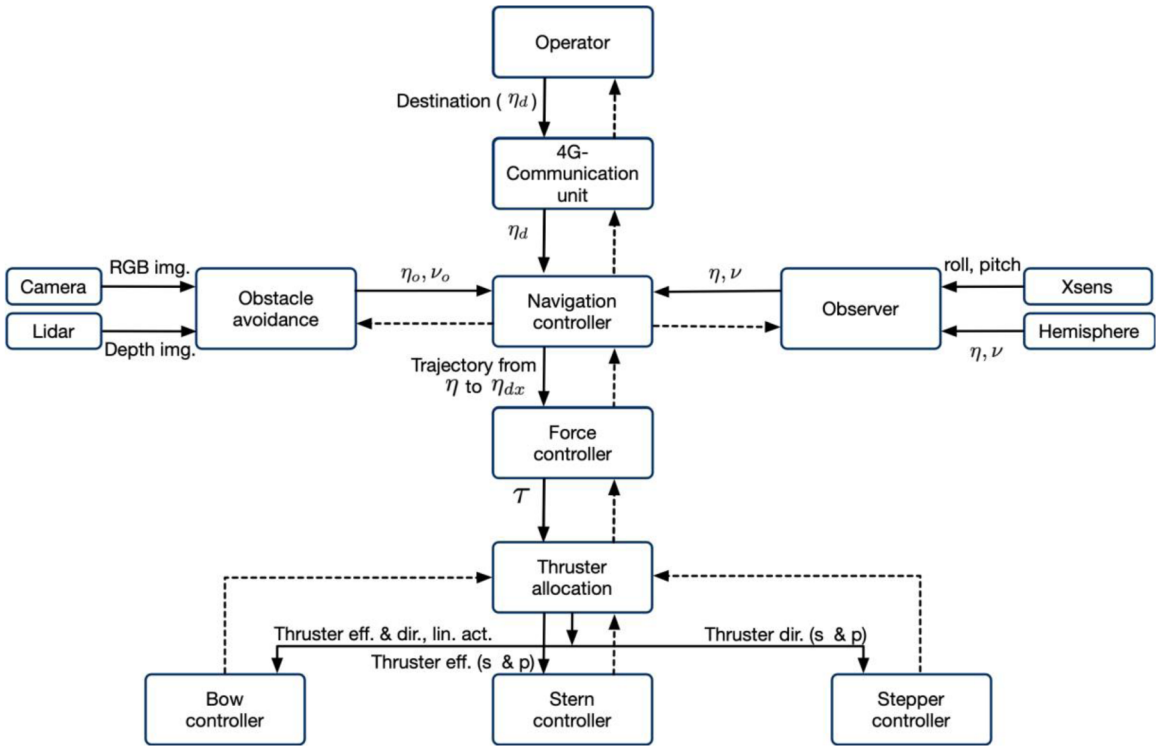


Figure C.1: The control structure of the ReVolt

Bibliography

[1] J. Sifakis, "Autonomous systems – an architectural characterization," eng, in *Lecture notes in computer science*, ser. Lecture Notes in Computer Science, vol. 11665, Cham: Springer International Publishing, 2019, pp. 388–410.

[2] *Autonomous all-electric passenger ferries for urban water transport (autoferry)*, https://www.ntnu.edu/autoferry, NTNU, (Accessed: 24.03.22).

[3] L. Franceschi-Bicchierai, *How a hacker controlled dozens of teslas using a flaw in third-party app*, https://www.vice.com/en/article/akv7z5/how-a-hacker-controlled-dozens-of-teslas-using-a-flaw-in-third-party-app, (Accessed: 09.06.22).

[4] A. Felski and K. Zwolak, "The ocean-going autonomous ship - Challenges and threats," eng, vol. 8, no. 1, p. 41, 2020.

[5] *Sfi autoship*, https://www.ntnu.edu/sfi-autoship, (Accessed: 16.06.22).

[6] *AUTOSHIP - Autonomous Shipping Initiative for European Waters*, https://www.autoship-project.eu/, (Accessed: 24.03.22).

[7] *The NTNU Digital Transformation initiative*, https://www.ntnu.edu/digital-transformation, (accessed: 24.03.22).

[8] H. E. Almaas, *Økninig i anntall ladestasjoner for ferjer og anlegg for landstrøm*, https://www.ssb.no/transport-og-reiseliv/sjotransport/artikler/okning-i-antall-ladestasjoner-for-ferjer-og-anlegg-for-landstrom, (accessed: 24.03.22).

[9] E. Eide, "Milliampère – norges første førerløseferje," in *Sjøsikkerhetskonferansen 2018*, 2018.

[10] *Google earth*, https://earth.google.com/web/.

[11] A. Amro, V. Gkioulos, and S. Katsikas, "Communication architecture for autonomous passenger ship," English, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2021.

[12] N. O. Servie, *What is lidar?* https://oceanservice.noaa.gov/facts/lidar.html, (Accessed: 09.06.22).

[13] *DNV class guidelines: Autonomous and remotely operated ships*, (Accessed: 24.03.22).

[14] *International Maritime Organization*, https://www.imo.org/, (accessed: 30.03.22).

[15] N. Leveson, "Safety and Security Are Two Sides of the Same Coin," eng, *The coupling of safety and security : exploring interrelations in theory and practice*, SpringerBriefs in Safety Management, pp. 17–27, 2020.

[16] "IEC 62443-1-1, Industrial communication networks - Network and system security - Part 1-1: Terminology, concepts and models," eng, International Electrotechnical Commission, Geneva, CH, Standard NEK IEC TS 62443-1-1:2009, 2009.

[17] M. Veale and I. Brown, "Cybersecurity," eng, *Internet policy review*, vol. 9, no. 4, pp. 1–22, 2020.

[18] International Electrotechnical Commission, *Electropedia: The world's online electrotechnical vocabulary*, https://www.electropedia.org/, (Accessed: 16.06.22).

[19] Cisco, *What is cybersecurity?* https://www.cisco.com/c/en/us/products/security/what-is-cybersecurity.html, (Accessed: 16.06.22).

[20] T. Salater, *NEK IEC 62443 – en bærebjelke for cybersikkerhet*, https://www.nek.no/nek-iec-62443-en-baerebjelke-for-cybersikkerhet/, (Accessed: 07.04.22).

[21] A. Myhrvold, chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ptil.no/contentassets/11851dc03a84473e8299a2d80e656356/barriers-memorandum-2017-eng.pdf, (Accessed: 16.06.22).

[22] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, *Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains*. 2011, pp. 113–125.

[23] M. J. Assante and R. M. Lee, *The Industrial Control System Cyber Kill Chain*, https://sansorg.egnyte.com/dl/HHa9fCekmc, (Accessed: 10.03.22).

[24] R. Grimmick. "What is C2? Command and Control Infrastructure Explained." (). (Accessed: 14.03.22).

[25] Septentrio, *What is spoofing and how to ensure GPS security?* https://www.septentrio.com/en/learn-more/insights/what-spoofing-and-how-ensure-gps-security, (Accessed: 18.06.22).

[26] B. Strom P, A. Applebaum, D. Miller P, K. Nickels C, A. Pennington G, and C. Thomas B, *MITRE ATT&CK: Design and Philosophy*, https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf, 2020, (Accessed: 01.04.22).

[27] O. Alexander, M. Belisle, and J. Steele, *MITRE ATTACK for Industrial Control Systems: Design and Philosophy*, 2020, (Accessed: 31.03.22).

[28] *ATTACK for Industrial Control Systems*, https://collaborate.mitre.org/attackics/index.php/Main_Page, (Accessed: 01.04.22).

[29] E. Lisova, I. Sljivo, and A. Causevic, "Safety and Security Co-Analyses: A Systematic Literature Review," eng, *IEEE systems journal*, vol. 13, no. 3, pp. 2189–2200, 2019.

[30] N. H. Carreras Guzman, J. Zhang, J. Xie, and J. A. Glomsrud, "A Comparative Study of STPA-Extension and the UFoI-E Method for Safety and Security Co-analysis," English, *Reliability Engineering and System Safety*, vol. 211, 2021.

[31] N. H. Carreras Guzman, M. Wied, I. Kozine, and M. A. Lundteigen, "Conceptualizing the key features of cyber-physical systems in a multi-layered representation for safety and security analysis," eng, *Systems engineering*, vol. 23, no. 2, pp. 189–210, 2020.

[32] N. H. Carreras Guzman, I. Kozine, and M. A. Lundteigen, "An integrated safety and security analysis for cyber-physical harm scenarios," eng, *Safety science*, vol. 144, pp. 105–458, 2021.

[33] N. H. Carreras Guzman, D. Kwame Minde Kufoalor, I. Kozine, and M. A. Lundteigen, "Combined safety and security risk analysis using the UFoI-E method: A case study of an autonomous surface vessel," English, Hannover, Germany, 2020, pp. 4099–4106.

[34] M. Rausand, *Risk Assessment: Theory, Methods, and Applications*, eng, ser. Statistics in practice. Somerset, 2011.

[35] *Cyber physical systems : From theory to practice*, eng, Boca Raton, 2016.

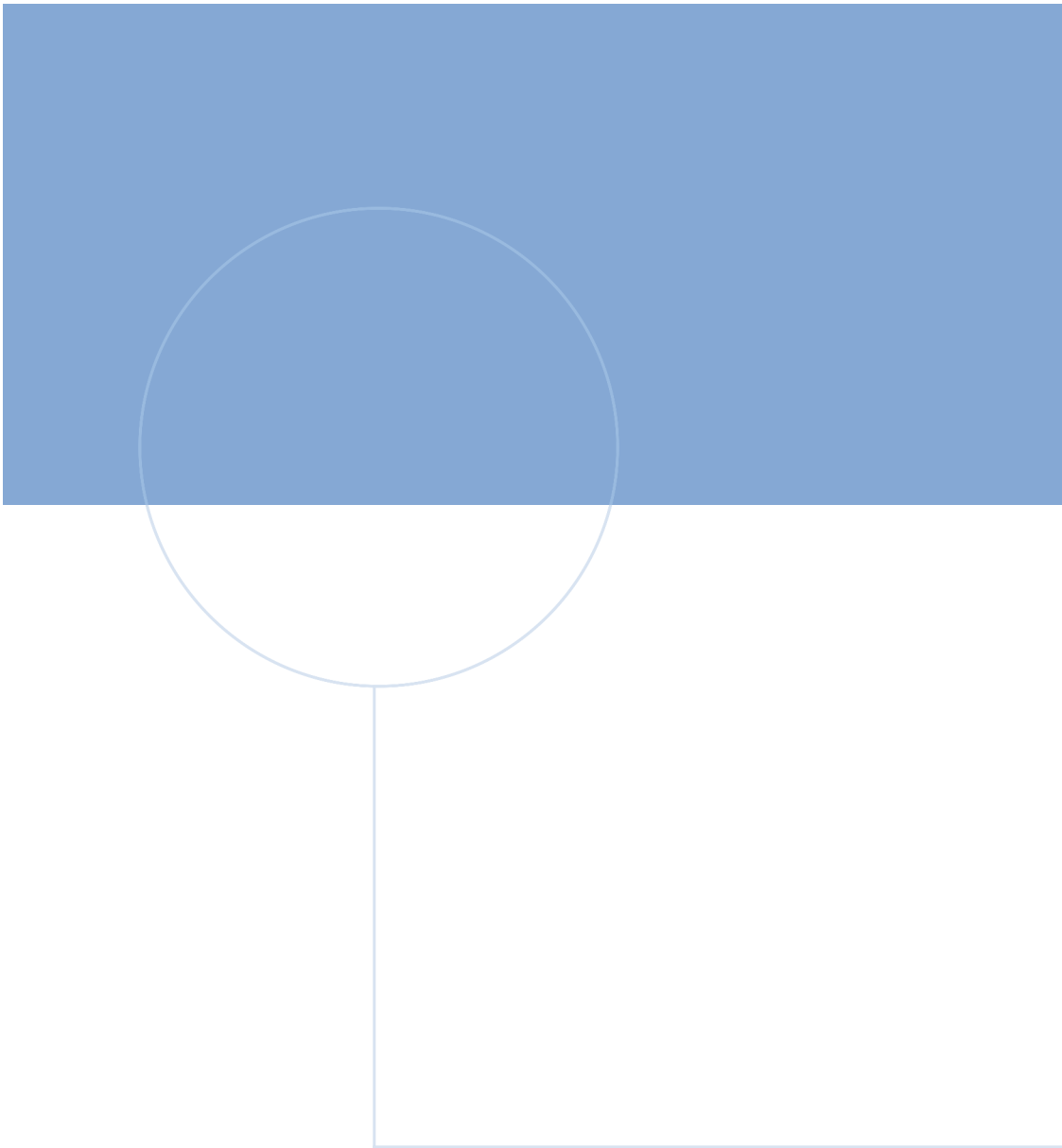
[36] N. G. Leveson and J. P. Thomas. "STPA HANDBOOK." (2018). (Accessed: 16.06.22).

[37] N. G. Leveson, *Engineering a Safer World : Systems Thinking Applied to Safety*, eng, ser. Engineering Systems. Cambridge: The MIT Press, 2012.

[38] W. Young and N. Leveson, "Systems thinking for safety and security," eng, in *Proceedings of the 29th Annual Computer Security Applications Conference*, ser. ACSAC '13, ACM, 2013, pp. 1–8.

[39] J. A. Glomsrud and J. Xie, "A structured STPA safety and security co-analysis framework for autonomous ships," English, Hannover, Germany, 2020, pp. 38–45.

[40] DNV, *The ReVolt*, `https://www.dnv.com/technology-innovation/revolt/`, (Accessed: 16.03.22).



NTNU

Norwegian University of
Science and Technology