A Comparative Study of STPA-Extension and the UFoI-E Method for Safety and Security Co-analysis

Carreras Guzman, Nelson Humberto; Zhang, Jin; Xie, Jing; Glomsrud, Jon Arne

Published in:
Reliability Engineering and System Safety

Link to article, DOI:
10.1016/j.ress.2021.107633

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Reliability Engineering and System Safety 211 (2021) 107633

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

A Comparative Study of STPA-Extension and the UFoI-E Method for Safety and Security Co-analysis

Nelson H. Carreras Guzman [a,b,*], Jin Zhang [a,c], Jing Xie [d], Jon Arne Glomsrud [d]

[a] Engineering Systems Design Section, Technical University of Denmark (DTU), Kgs, Lyngby, 2800, Denmark
[b] Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, 7034, Norway
[c] Department of Computer Science, Norwegian University of Science and Technology (NTNU), Trondheim, 7034, Norway
[d] Group Technology and Research, DNV GL, Høvik, 1363, Norway

ARTICLE INFO

Keywords:
Safety and security
comparative study
risk identification
cyber-physical systems (CPSs), autonomous ship

ABSTRACT

Emerging challenges in cyber-physical systems (CPSs) have been encouraging the development of safety and security co-analysis methods. These methods aim at mitigating the new risks associated with the convergence of safety-related systemic flaws and security-related cyber-attacks that have led to major losses in CPSs. Although several studies have reviewed existing safety and security co-analysis methods, only a few empirical studies have attempted to compare their strengths and limitations to guide risk analysis in practice. This paper bridges the gap between two novel safety and security co-analysis methods and their practical implementations. Namely, this paper compares a novel extension of the System-Theoretic Process Analysis (STPA-Extension) and the Uncontrolled Flows of Information and Energy (UFoI-E) method through a common case study. In our case study, the CPS under analysis is a conceptual autonomous ship. We conducted our comparative study as two independent teams to guarantee that the implementation of one method did not influence the other method. Furthermore, we developed a comparative framework that evaluates the relative completeness and the effort required in each analysis. Finally, we propose a tailored combination of these methods, exploiting their unique strengths to achieve more complete and cost-effective risk analysis results.

1. Introduction

Increasingly, cyber-physical systems (CPSs) integrate novel information technologies and higher levels of automation into physical world operations. However, as CPSs are becoming both safety-critical and security-critical in real-world applications, one of the main challenges at present and for the future of risk science is the integration of safety and security analysis [50]. Indeed, safety-related systemic flaws and security-related cyber-attacks have overlapped in their contribution to recent hazardous events in CPSs [9]. Examples include major damages and losses in industrial control systems, autonomous vehicles, medical devices, among others [21,47].

Traditionally, the domain of safety analysis was bounded to accidental or unintentional risks, whereas the domain of security analysis focuses on intentional sources of risk [1,35]. More recently, researchers have proposed and reviewed different methods that integrate safety and security analysis into a co-analysis framework. In the literature, comprehensive surveys have assessed the distinguishing features of many of these novel safety and security co-analysis methods, providing theoretical classifications and insights about their capabilities [10,23,25,33,34].

However, to the best of our knowledge, there is no sufficient empirical studies to assess the benefits and limitations of applying these co-analysis methods in real systems. The novelty of these methods results in little evidence that can demonstrate their applicability and usefulness to mitigate the emerging risks in CPSs. Moreover, the complexity, openness and novelty of autonomous systems such as self-driving vehicles, industrial robots, and autonomous ships carry with them unprecedented cyber risks, which likely cannot be handled by traditional safety or security risk analysis methods. We need to explore whether those novel safety and security co-analysis methods can properly identify safety hazards and security threats and effectively mitigate them.

In this paper, we compare two newly developed safety and security co-analysis methods in a common case study to bridge the gap between the theoretical methods and their practical implementation. Namely, we

* Corresponding author: Akademivej, Building 358, room 920, 2800 Kgs. Lyngby. Denmark
E-mail address: nelca@dtu.dk (N.H. Carreras Guzman).

https://doi.org/10.1016/j.ress.2021.107633
Received 31 August 2020; Received in revised form 7 December 2020; Accepted 11 March 2021
Available online 13 March 2021

compare the Uncontrolled Flows of Information and Energy (UFoI-E) method [6] and a novel extension of the System-Theoretic Process Analysis (STPA-Extension) [19]. The system used in the comparative study is a conceptual ship with a revolutionary concept for unmanned, zero-emission, shortsea shipping, which is called the ReVolt[1].

Considering the comprehensive survey of the literature in Kavallieratos et al. [23], the UFoI-E method and STPA-related methods are two suitable alternatives for safety and security co-analysis. After their theoretical assessment, the authors of this survey concluded that both methods have a systematic and structured process that is scalable for different levels of complexity throughout the system lifecycle of CPSs. This assessment supports the claim that these two methods are notable candidates for an empirical comparison, where we can test their applicability and usefulness within the context of the safety and security co-analysis of a real CPS.

Our primary contribution in this work is three-fold. Firstly, we exemplify the concrete implementation process (stepwise) of applying UFoI-E and STPA-Extension to the risk analysis of an autonomous ship. The insights from this comparative study can be further developed into industrial guidance on risk assessment. Secondly, we leverage the analysis results into a novel comparative framework and evaluate the two methods from two fundamental aspects, i.e., completeness of the analysis results and effort spent on analysis. Such comparison results provide a solid argument for risk analysts who need to make a trade-off between analysis scope and analysis effort cost. Lastly, we propose potential improvements to combine the strengths of the two methods and enhance them to achieve a more comprehensive and cost-effective risk analysis.

This paper is organized as follows. Section 2 summarizes the features of UFoI-E and STPA-Extension and reviews the literature in comparative studies of safety and security co-analysis methods. Section 3 illustrates the design of our case study and presents the ReVolt as the system under analysis. Section 4 describes and compares the results obtained from UFoI-E and STPA-Extension, demonstrating the capabilities of each method. Section 5 discusses the comparison of results and suggests strategies to improve the overall safety and security co-analysis. Section 6 concludes the paper.

2. Related work

In this section, we briefly introduce UFoI-E and STPA-Extension to provide the necessary background knowledge for the following content of the paper. Then, we review the state-of-the-art on comparing safety and security co-analysis methods in practice.

2.1. The Uncontrolled Flows of Information and Energy (UFoI-E) method

The Uncontrolled Flows of Information and Energy (UFoI-E) method supports risk analysts in the identification and mitigation of harm scenarios in CPSs. This method provides a systems engineering approach to model complex risk scenarios in CPSs, identify safety and security risk sources, map propagation effects across the layers of the system and finally provide a layers of protection strategy to mitigate the risks.

As shown in Fig. 1, the UFoI-E method consists of three main constituents:

(a) UFoI-E causality concept

The first constituent of the UFoI-E method is the UFoI-E causality concept. The UFoI-E causality concept is a model of causation that considers safety and security risks into an integrated model. Accident causation models are highly regarded as the fundamental theories supporting all safety analysis methods [20,29]. However, these theoretical foundations of safety analysis were separated from the field of security engineering [3,20,28,38]. Therefore, for the first time in the domain of safety analysis, the UFoI-E causality concept provides an accident causation model that integrates safety and security engineering with the incorporation of physical and cybersecurity threats as risk sources [5,7].

The UFoI-E causality concept is an extension of the Uncontrolled Flows of Energy (UFoE) model [17,36]. In the UFoE model, an undesired release of energy – e.g. potential, thermal, or electrical – or a release of toxic materials can lead to physical harm to valuable entities. To avoid or mitigate harm, designers can allocate safety barriers to separate the UFoE from the valuable entities. The UFoI-E causality concept extends this UFoE model to include the ways that Uncontrolled Flows of Information (UFoI) could also kill humans and cause physical damages. Cases of UFoI stress how software flaws, network design errors and novel cyber-physical attacks threaten the safety of CPSs and could lead to human fatalities, asset damages and environmental impacts. Consequently, a combination of safety barriers and security barriers is needed to avoid or mitigate harm [7].

As the first constituent of the UFoI-E method, the UFoI-E causality concept is the only constituent that is not an operational tool of the method. Instead, this constituent is the fundamental theory that supports the other two practical constituents.

(b) CPS master diagram

The second constituent of the UFoI-E method is the CPS master diagram. The UFoI-E method emphasizes the need to represent and visualise the CPS under analysis using a comprehensive framework. In the words of P. L. Clemens, “we never analyse a system – we analyse only a conceptual model of a system” [38].

For this purpose, this method provides the CPS master diagram, a generic representation of CPSs that teams of analysts can use as an initial template to translate their particular system specifications into a tailored diagram. The CPS master diagram establishes a common terminology and a system model to integrate the knowledge of multidisciplinary risk analysis teams. In general, the CPS master diagram conceptualises CPSs as a system of three interacting layers: the cyber layer (CL), the cyber-physical layer (CPL), and the physical layer (PL).

In a nutshell, the PL is the domain of the system that influences the physical world. This PL includes the energy flows and material processes monitored by humans via manual and analog interfaces. The CPL is the domain of the system where networked control systems perform real-time and autonomous actions to influence the PL. This CPL includes sensors, programmable controllers and actuators that transmit information flows via real-time communication networks. The CL is the domain of the system that monitors the rest of the CPS as supervisory control technologies and processes. This CL includes computer networks and humans that monitor information flows coming from the PL and the CPL. Finally, all these layers also interact with their cyber and physical environments, exchanging flows of information and energy. For a comprehensive description of the CPS master diagram, see Carreras Guzman, Wied, et al. [9].

As the second constituent of the UFoI-E method, the CPS master diagram stresses the need to agree on a common model of the CPS before conducting the identification of safety and security risk scenarios. Consequently, a generic CPS master diagram can be used to create a tailored diagram of the specific CPS under analysis.

(c) CyPHASS

The third and final constituent of the UFoI-E method is the Cyber-Physical Harm Analysis for Safety and Security (CyPHASS). CyPHASS is a harm scenario builder to assist a systematic identification of safety and security risk scenarios using an extended bowtie model. Overall, CyPHASS is a toolkit composed of two main parts: (1) an ontology of scenarios and (2) a database of checklists and guidewords.

(1) Ontology of scenarios: The ontology of scenarios in CyPHASS is an extended bowtie model that illustrates a comprehensive set of generic paths that could lead to unsafe consequences in CPSs.

[1] The ReVolt is developed by DNV GL: https://www.dnvgl.com/technology-innovation/revolt/index.html

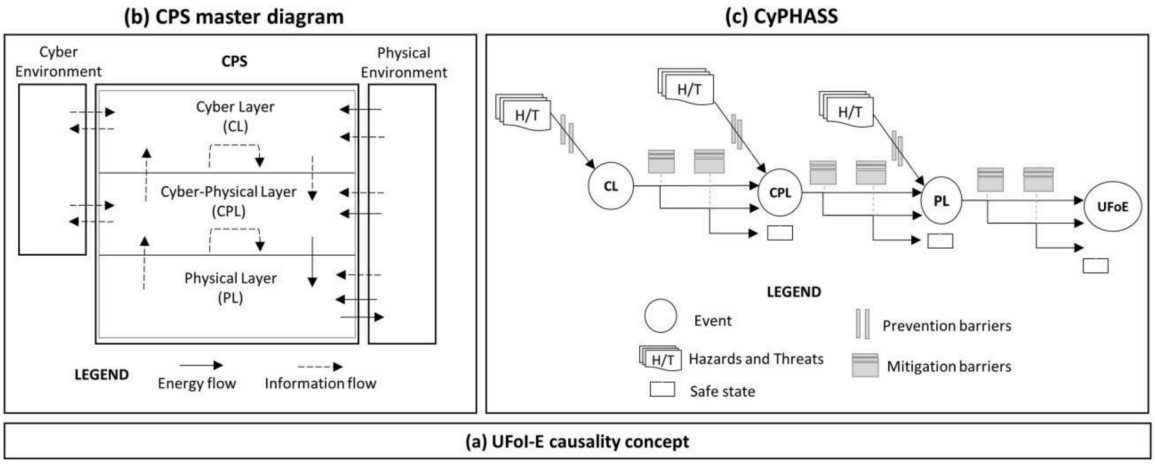

Fig. 1. Constituents of the UFoI-E method: (a) UFoI-E causality concept; (b) CPS master diagram, (c) CyPHASS.

Following the conventions of bowtie models [13], the ontology of scenarios describes an extended bowtie composed of three consecutive top events, each of them associated to their causes to the left and their potential consequences to the right. As depicted in Fig. 2, various causes and propagation effects could cascade across the layers of the CPS master diagram and lead to complex scenarios in a series of stages. This ontology of scenarios is a generalisation that considers the experience from a wide set of hazardous events in CPSs, where accidents and cyber-attacks have demonstrated how risk sources affecting one layer of the CPS have cascaded throughout the connected layers and produced catastrophic damages. Indeed, these cascades across layers have occurred in diverse CPS applications including industrial plants, driverless vehicles, medical devices, among others [6,21,48].

(2) Database of checklists and guidewords: CyPHASS includes a database of checklists and guidewords for each stage of the scenario. This database is a knowledge repository of lessons learned and expert knowledge to assist risk analysis teams [6]. As shown in Fig. 2, for each component of the extended bowtie in the ontology of scenarios, a list of checklists and guidewords is available to be used as supporting material. As a prototype tool, this database of checklists and guidewords is available as an open-source material for risk analysts [4]. In combination, the CyPHASS ontology of scenarios and its database of checklists and guidewords provide a tool to identify harm scenarios and to recommend barriers to prevent and mitigate these scenarios.

2.1.1. *Application process of the UFoI-E method*

After a team of analysts translate their system specifications into a tailored CPS master diagram, the risk identification with CyPHASS is a stepwise process. Previous works have illustrated how to develop tailored CPS master diagrams for safety and security analysis [7–9]. In the following paragraphs, we will focus on the systematic application of CyPHASS, which is a novel development of the UFoI-E method [6].

The risk identification process with CyPHASS addresses how unsafe consequences can arise from different risk sources targeting – either unintentionally or intentionally - the different layers of the CPS. In line with the Society for Risk Analysis Glossary, the UFoI-E method uses the term “hazard” for unintentional or natural sources of risks associated to safety, whereas the term “threat” corresponds to intentional sources of risk associated to security [1,2]. In the CyPHASS ontology, hazards and threats (H/T) are the initial risk sources in the scenarios.

In a systematic way, the process of scenario identification with CyPHASS consists on linking a set of ultimate safety consequences all the way back to their initial risk sources. In CyPHASS, the ultimate safety consequences occur at the physical layer (PL) and the physical environment of the CPS master diagram. At these layers, the physical entities of the system – humans, assets and the natural ecosystem - can suffer

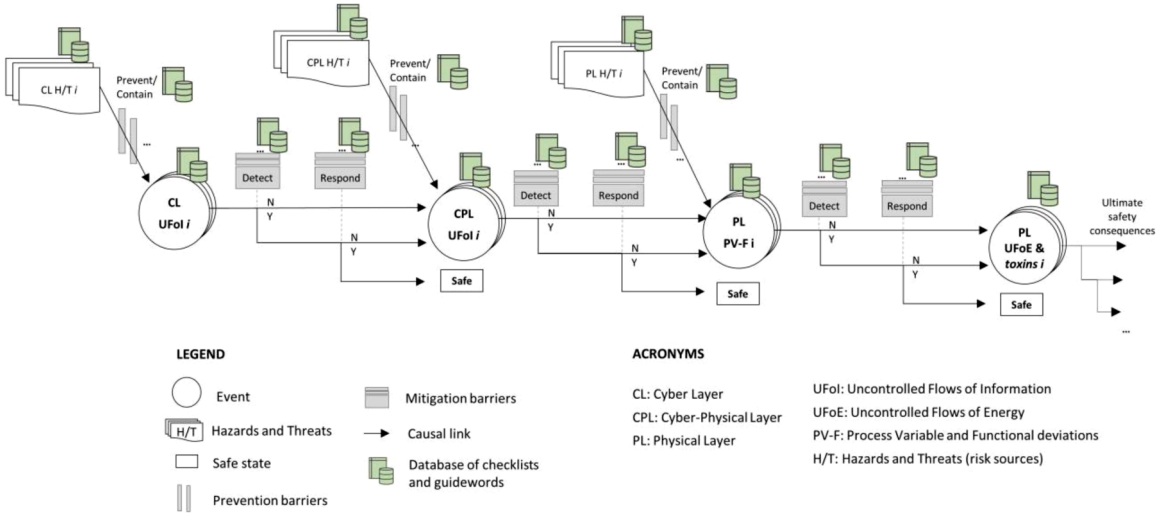

Fig. 2. The CyPHASS ontology of scenarios as an extended bowtie model, adapted from Carreras Guzman, Kozine et al. [6].

injuries and damages as ultimate safety consequences. In CyPHASS, the direct causes of these ultimate safety consequences are uncontrolled flows of energy (UFoE) and toxins occurring at the PL. Examples include uncontrolled kinetic or thermal energies, releases of toxic and radioactive substances, among others.

In CyPHASS, an UFoE is the result linked to one – or a combination of - process variable and functional deviations (PV-F) arising at the PL. To discover these potential deviations, the CyPHASS database recommends the use of hazard and operability (HAZOP) guidewords [14,45] and functional failure checklists [15]. This integration of HAZOP guidewords and functional failure checklist offers an comprehensive toolkit to analyse the PL. Examples include deviations associated to actions of human operators at the PL, deviations in mechanical and electrical parameters at the PL, among others [6].

Starting from the PV-F deviations to the left, the causes in the bowtie ontology can be separated into two categories. The first category is a direct risk source (H/T) threatening the system layer under analysis. The second category is a propagation effect from the adjacent layer of the CPS. In CyPHASS, these propagation effects can arise as intermediate events that cascade across the layers of the system. Moreover, a concurrent presence of direct risk sources and propagation effects can also lead to the discovery of complex scenarios.

As causes of PV-F deviations at the physical layer, the propagation effect can be the result of an uncontrolled flow of information (UFoI) occurring at the cyber-physical layer (CPL). Examples of UFoI include corrupted, delayed or missing information flows. Specifically, these UFoI at the CPL are deviations associated to real-time control subsystems - i.e. sensors, programmable controllers, actuators and the real-time communication networks that transmit their information flows -. As with the PL, the CPL is exposed to direct risk sources as H/T as well as to additional propagation effects. In this final case, the propagation effects arise from the cyber layer (CL).

In CyPHASS, UFoI occurring at the CL are the final and most distant stage associated to the unsafe safety consequences. Nevertheless, recent catastrophic events in industrial plants, driverless vehicles and medical devices have shown how cases of UFoI at the CL can propagate across all the layers of the CPS and lead to unsafe safety consequences [6,21,48]. Notable examples are the Stuxnet attack [27] and the TRITON attack [22]. The risk sources associated to these particular CL UFoI include unintentional supervisory errors as well as intentional cyber-attacks.

As referred by Carreras Guzman, Kozine et al. [6] and illustrated in Fig. 3, CyPHASS describes a stepwise process to perform risk identification backwards, i.e. starting from the ultimate consequences and discovering all the stages of the scenarios towards the initial risk sources. The steps can be summarized as follows:

"**Step 1**: Identify the cases of UFoE that could lead to ultimate safety consequences
Step 2: For each UFoE, identify the causes as PL PV-F deviations
Step 2.1: For each PL PV-F deviation, identify and recommend detection and response barriers
Step 3: For each PL PV-F deviation, identify causes as physical hazards and threats (H/T)
Step 3.1: For each physical H/T, identify and recommend prevention barriers
Step 4: For each PL PV-F deviation, identify causes as CPL UFoI
Step 4.1: For each CPL UFoI, identify and recommend detection and response barriers
Step 5: For each CPL UFoI, identify causes as cyber and physical H/T
Step 5.1: For each cyber and physical H/T, identify and recommend prevention barriers
Step 6: For each CPL UFoI, identify causes as CL UFoI
Step 6.1: For each CL UFoI, identify and recommend detection and response barriers
Step 7: For each CL UFoI, identify causes as cyber and physical H/T
Step 7.1: For each cyber and physical H/T, identify and recommend prevention barriers" [6]

Notice that in CyPHASS, an event tree with mitigation barriers follows to the right of each intermediate event. These mitigation barriers are detection and response countermeasures that aim at avoiding the propagation of the scenario to reach a safe state. At the left of each intermediate event, a set of hazards or threats (H/T) can be direct causes for each intermediate event. To prevent the H/T, prevention barriers can eliminate or reduce the likelihood of the H/T.

In sum, CyPHASS suggests the allocation of sequential prevention and mitigation barriers acting as layers of protection throughout the stages of the scenarios. This layers of protection strategy is considered critical to ensure that, even if one safety or security barrier is breached, other barriers can be activated across the layers of the system. As an illustrative case, even if a malware – e.g. Stuxnet or TRITON – infects a computer network at the CL and attempts to propagate to the control network at the CPL, additional barriers can still detect and respond at the CPL or at the PL to contain or mitigate damages. These barriers include technical components and system architecture arrangements as well as human operations and organisational strategies.

At each step of the stepwise process in CyPHASS, the analysts can use a database of generic checklists built from lessons learned and expert knowledge of CPSs. This overall database is available as a software prototype in common spreadsheets format and is shared as an open-source material [4]. For each case in the database, the analysts can

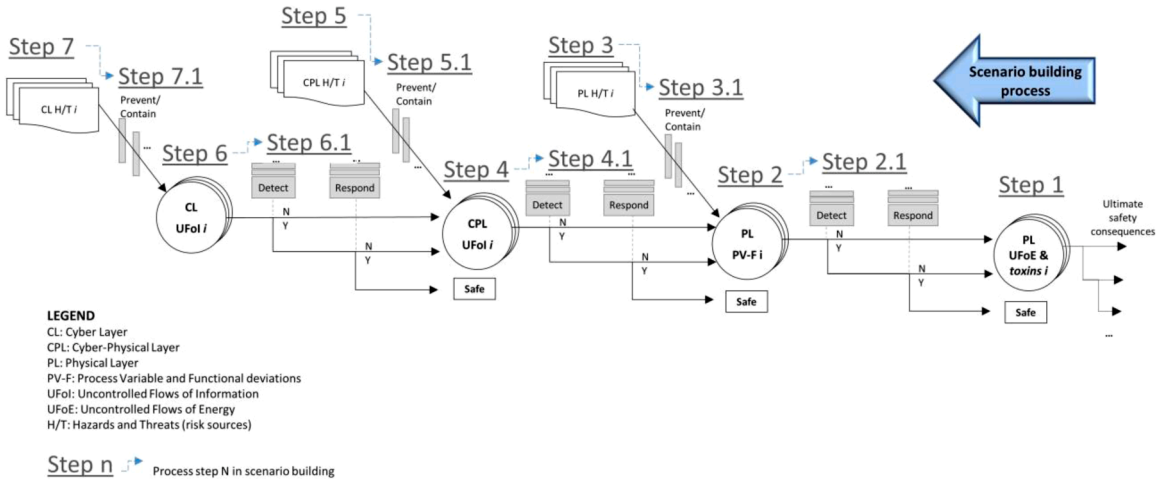

Fig. 3. CyPHASS extended bowtie model in analysis steps, adapted from Carreras Guzman, Kozine et al. [6].

visualise their CPS master diagram, ask the question *"Is this case possible in my system?"* and link the connections between the sequential stages of their scenarios in the CyPHASS ontology.

2.2. STPA safety and security co-analysis framework

The Systems-Theoretic Process Analysis (STPA) was primarily developed as a safety hazard analysis technique based on the Systems-Theoretic Accident Model and Processes (STAMP) [30]. STPA is a top-down approach consisting of four fundamental steps [32] listed below:

- Step 1: Define the purpose of the analysis and establish the system engineering foundation
- Step 2: Model the control structure
- Step 3: Identify unsafe control actions
- Step 4: Identify causal scenarios

Later, STPA-Sec was proposed to extend the hazard analysis scope by incorporating systems security engineering into analysis as well [49]. STPA-Sec maintains the four fundamental steps unchanged while essentially introducing security threats/vulnerabilities identification into Step 4. More recently, several researchers have proposed extensions of STPA for safety and security co-analysis [16,24,39].

When applying the conventional STPA or STPA-Sec approaches to the analysis of novel CPSs, such as autonomous ships, the method faces several challenges. Firstly, Step 1 does not provide sufficient analysis artefacts to guide the system design of novel systems, which are typically in the exploration phase and not well defined. It is not straightforward to model the control structure in Step 2 with limited prior knowledge and experience of such systems. To bridge such a gap, a structured STPA safety and security co-analysis approach was proposed [19]. The structured co-analysis approach is depicted in Fig. 4 and is referred to in this paper as STPA-Extension.

As shown in Fig. 4, this co-analysis approach introduces *Functional Requirements,* i.e., sub-step number (6), into Step 1 to facilitate the development of the control structure in Step 2. Secondly, STPA-Sec does not explicitly consider the security-related losses, which could be equally critical as safety-related losses for many systems. To cope with this issue, the structured STPA co-analysis approach explicitly includes system-level security-related losses when identifying the system-level losses. It also differentiates the system-level security incidents from safety accidents to guide the identification of system-level hazards in Step 1. It further enhances the co-analysis in Step 4 where not only accidental and unintentional, but also intentional causes, are comprehensively analysed at component-level. In short, there are two separate aspects of security and they need to be separately included in the co-analysis process. These are (1) security-related causes can lead to safety-related losses, and (2) safety-related causes can lead to security-related losses.

2.3. Review of comparative studies of safety and security co-analysis methods

This subsection reviews empirical studies comparing risk analysis methods with a focus on the safety and security of CPSs. Based on this body of knowledge, we discuss a research gap that hinders the validity and replicability of these comparative studies.

Kriaa et al. [26] compared the Combined Harm Assessment of Safety and Security for Information Systems (CHASSIS) [37] and the Boolean logic Driven Markov Processes (BDMP) [35] in a case study. The case study was a CPS previously modelled with BDMP. During the comparison, the BDMP model was changed to match the CHASSIS model. At this point, the two methods influenced each other during the analysis to have "harmonizing" models of system representation. This comparison also highlighted the strengths of both methods and showed the possibility to combine them.

Schmittner et al. [40] compared CHASSIS and the Failure Modes, Vulnerabilities and Effects Analysis (FMVEA) for an automotive CPS. This study showed that the two methods do not have many overlapping failures, as the FMVEA found more component-based failures and the CHASSIS found more software and cyber threats. This study also stated that the CHASSIS method analyses safety and security as two separate parameters, which is not ideal as they could affect each other. A unified

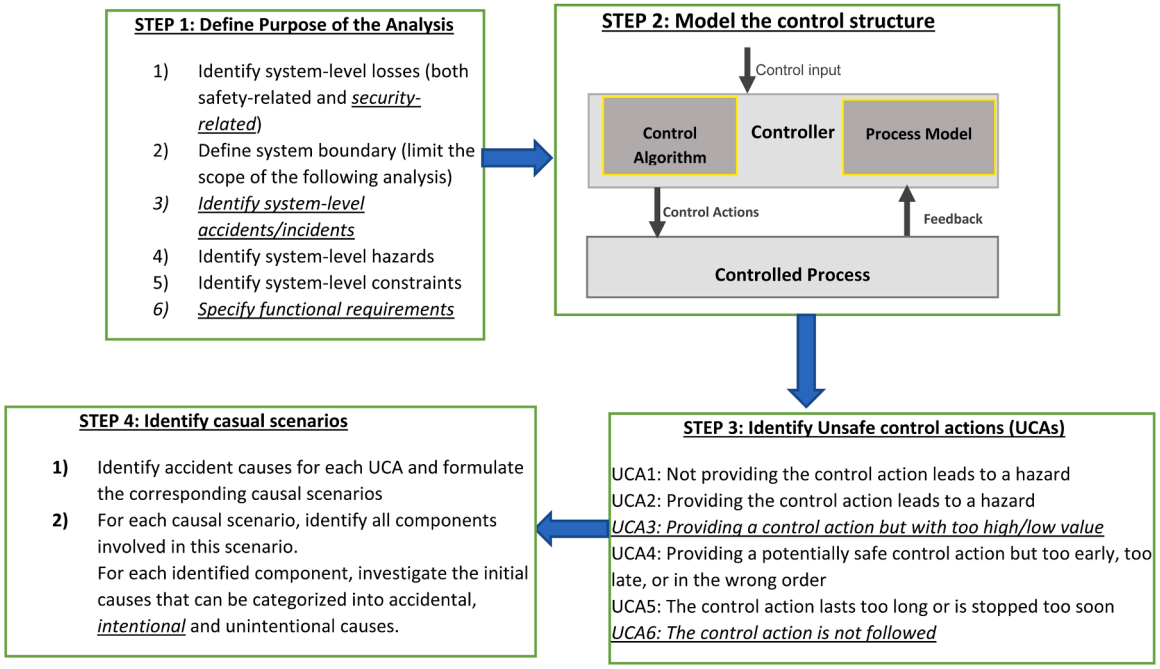

Fig. 4. A Structured STPA Safety and Security Co-analysis framework (STPA-Extension).
(Bold texts: describe the conventional STPA steps in details. Italics texts with underline: highlight the structured STPA co-analysis improvements to concretize the analysis steps).

risk rating for threats and failures that influences security would be a necessary improvement for both methods. This paper did not evaluate the amount of time and effort needed for both analyses.

Sulaman et al. [43] compared the STPA and the Failure Modes and Effects Analysis (FMEA) method by applying STPA hazards categories to the FMEA failures found in their case study. For comparison purposes, they used an assessment criteria following the technology acceptance model (TAM) [12]. This study concluded that the difference between the exposed hazards and failures is little to almost indifferent between the methods. However, their result showed that FMEA provides more specific failures, whereas STPA provides more detailed causes. According to the analysis, the difficulty of the two methods seemed to be very similar. When comparing the effort to use the two methods, they only made a rough estimation of the time used to determine if there was a relationship between effort, time, and outcome. The main limitation of their comparison is that they categorized the items in the STPA categories. Therefore, the hazards found from the STPA analysis look best illustratively, since the categories were made for the STPA method.

Wei et al. [46] used the STPA-Sec method to identify threats to a conceptual mobility-as-a-service fleet of autonomous vehicles. After obtaining the STPA-Sec results, they performed a CHASSIS analysis to complement their STPA results rather than to compare the two methods. The two analyses were not performed on an equal base, which led to different results. As the same authors applied the two methods, they gained knowledge in the process of applying the different methods. However, as each method has a different objective, it is hard to tell to what degree this gained knowledge affected the outcome.

The reviewed literature mainly compared the complexity and the outcome obtained from the different safety and security co-analysis methods. However, we identify four specific issues that hinder the validity and replicability of these previous studies:

1 Some comparative studies do not take sufficient measures to avoid the problem of accumulation of knowledge. This problem arises when a single team of analysts performs the first analysis with one safety and security co-analysis method and then performs the second analysis with the second method. For the second analysis, the single team is carrying the knowledge gained during the first analysis and influencing the results obtained in the second analysis.
2 Most comparative studies fail to report the level of knowledge of the analyst teams. We argue that, when two teams of analysts have differences in their (a) level of experience as analysts and (b) knowledge about the system under analysis; the results of the comparison will favour the safety and security co-analysis method that was used by the more experienced and knowledgeable team of analysts.
3 Most comparative studies give little or no indication about the effort required to obtain the results with each safety and security co-analysis method. Arguably, if more time and resources are dedicated for one method, this additional effort will favour the method in question.
4 Finally, some comparative studies only consider how one safety and security co-analysis method complements the results obtained from a previous analysis with a different method. These studies not only suffer from the issue of accumulation of knowledge, but also tend to reflect an a priori preference towards one of the methods. We argue that the goal of this type of studies is not to compare the two methods, but only to use some insights to refine the preferred method.

In the following section, we address these issues by presenting our research methodology for comparison of safety and security co-analysis methods. Accordingly, the goal is to present a framework to facilitate the validity and replicability of these comparative studies.

3. A framework to compare safety and security co-analysis methods

3.1. Research Methodology

The research objective of this empirical study is to evaluate the qualitative strengths and limitations of two safety and security co-analysis methods in terms of (1) completeness of the results and (2) effort required. Furthermore, this paper aims at identifying potential ways to combine these two methods effectively to overcome their respective limitations.

According to Creswell [11], case studies are research methods that are practical for the evaluation of qualitative data. As previously illustrated in Section 2.3, other comparative studies of safety and security co-analysis methods have used case study research. Therefore, we designed a case study as a suitable research method to achieve our research objective, overcoming some limitations identified in other comparative studies.

In particular, our case study is a prototypical implementation with feedback. Our prototypical implementation is the analysis of one common system using two different safety and security co-analysis methods. As shown in Fig. 5, the common system under analysis is the ReVolt autonomous vessel prototype, and the safety and security co-analysis methods are the UFoI-E method and STPA-Extension. After performing the analyses, the feedback is the comparison of the results obtained using each safety and security co-analysis method.

In our case study design, we designated two independent teams. Two authors of this paper composed Team 1, and the other two authors of this paper composed Team 2. Each team used only one safety and security co-analysis method, namely, Team 1 deployed the UFoI-E method and Team 2 used STPA-Extension. To avoid the issue of accumulation of knowledge, the two teams performed the analyses independently.

In terms of knowledge of the system, from the initial stage onwards, the teams shared a common documentation of the system specifications and a limited set of preliminary analyses performed for previous versions of the ReVolt. Regarding the knowledge of the other team's method and results, the teams conducted workshops to train each other in the generic features of their respective safety and security co-analysis methods. During the analysis, the teams only communicated and exchanged ideas to agree on the ways to represent their results in a comparable framework. This comparable framework implied the need to specify a common scope of the analysis and a similar level of abstraction.

To set a common scope of the analysis, the teams agreed to consider a unique ultimate consequence as system loss. This unique system loss is a collision of the ReVolt vessel while operating in autonomous mode. This scope of the analysis excludes the issues related to the other three modes of operation of the ReVolt (see Section 3.2). Moreover, the specification of a unique system loss emphasizes safety as the goal of the analysis, excluding other types of system losses such as financial losses, reputation losses, data privacy, among others.

To set a similar level of abstraction, the system specifications of the ReVolt as currently designed are a shared constant for both teams. These system specifications considered the hardware architecture and some

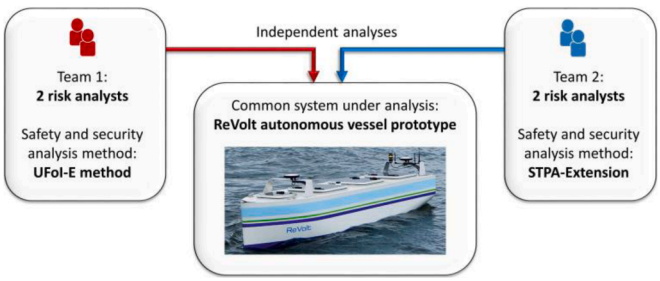

Fig. 5. Overview of our case study design.

technical specifications of the components as installed. For the software design and implementation, the level of abstraction was limited to the functional requirements of the control algorithms. Therefore, the analyses did not consider specific errors or vulnerabilities in the software code or the communication protocols. This level of abstraction is consistent with a systems perspective shared by both safety and security co-analysis methods deployed.

3.1.1. Evaluation criteria for the comparison

Completeness: To assess the quality and completeness of risk identification methods, Taylor [44] propose a theoretical measure of completeness. It defines completeness in terms of the *number of risks identified* by the method with respect to the *total number of risks* existing in the system under analysis. Therefore, as shown in Equation (1), the theoretical measure of completeness can be defined as a ratio.

$$\text{Completeness} = \frac{\text{Number of risks identified}}{\text{Total number of risks}} \quad (1)$$

If the ratio were equal to one, the risk identification would be "absolutely complete". However, the *total number of risks* is an objective benchmark that is usually not available for novel and complex systems. Therefore, one way to obtain a relative evaluation of completeness in risk identification is to compare several analyses of the same system [44].

In our comparative framework, the *number of risks* correspond to the *number of risk scenarios* identified by the safety and security co-analysis methods. In this paper, we define a risk scenario as a combination of conditions and events that can lead to a loss. As previously mentioned, in our case study we predefine a unique safety-related loss, namely, a collision of the ReVolt vessel while operating in autonomous mode. Therefore, different risk scenarios are different combinations of conditions and events – including common causes and sequential events – leading to a collision of the ReVolt. Considering the safety and security co-analysis of a CPS, the risk scenarios include accidental, unintentional and intentional causes related to cyber and physical parts of the system.

Subsequently, Fig. 6 illustrates our comparative framework for a relative evaluation of completeness in a Venn diagram. Each set of this Venn diagram corresponds to the total number of risks identified by each method. At their intersection, one can enumerate the risks identified by both methods independently. Conversely, the relative complements are the risks identified only by one method and not by the other. Finally, the union of the two sets corresponds to the relative benchmark of *total number of risks*.

In this comparative study, we identify two types of reasons that result in the relative complements in the Venn diagram – i.e. in the results obtained only by one method:

1 Team-specific: reasons associated with the teams using the method
2 Method-specific: generic reasons associated with the safety and security co-analysis methods as such.

These reasons serve to explain the differences obtained by the two methods. If the difference is team-specific, a proper application of the two methods can provide sufficient guidance to identify those risks missed by one of the teams. The team that missed those risks may have overlooked them due to different factors that are typical in risk analysis tasks (see Section 5.1). Conversely, if the difference is method-specific, it reveals a relative strength of the method that identified those specific risks. By identifying similarities and patterns between the method-specific results, we can propose potential improvements in which one method could complement the other in a combined analysis to obtain better results.

Effort required: To account for the background experience and knowledge about the ReVolt system, Table 1 explicitly shows these criteria distributed across the team members. This table includes the years of experience as safety and security analysts and the background knowledge about the ReVolt system at the beginning of the case study.

Table 1
Background experience and knowledge about the system before the comparative study.

Criteria	Team 1 (UFoI-E)		Team 2 (STPA-Extension)	
	Member 1	Member 2	Member 1	Member 2
Experience as a safety and security risk analyst (Years)	2	2	2	2
Background knowledge of the ReVolt system (Basic / Intermediate / Advanced)	Intermediate	Intermediate	Advanced	Intermediate

Basic: Minimum to null knowledge about the ReVolt system before the case study
Intermediate: Prior knowledge about the ReVolt system via public documents and scientific publications
Advanced: High knowledge of the ReVolt system through active involvement in design and/or operations

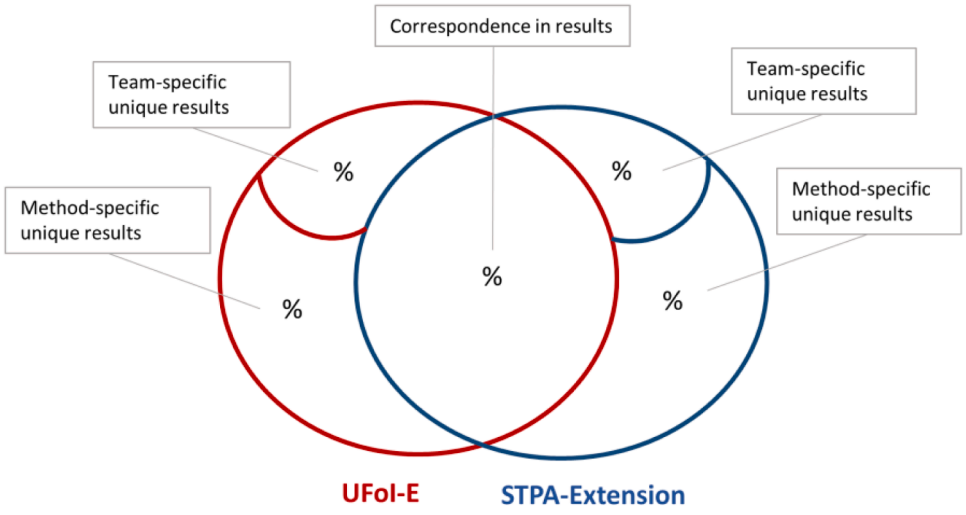

Fig. 6. Conceptual framework to compare the results as a relative evaluation of completeness.

After the teams concluded their independent analyses, they reported the total working hours used to prepare for and to perform their respective safety and security co-analysis. In conjunction, our comparative framework accounts for the total working hours used in each method and the level of knowledge of each team as the contributing factors leading to team-specific differences in the analysis results.

Furthermore, the working hours used in each method serve as an indicator of the cost or the resources needed to apply the method. In short, if both methods were to obtain the same results, the method that facilitates these results in less time would arguably be more cost-effective. However, this simplified reasoning assumes the same level of knowledge of the system and comparable expertise as a risk analyst, while it neglects the synergic interactions between the team members. Therefore, we provide this indicator of working hours alongside the information about the team members, including their knowledge of the system and the tools used for the analysis. We argue that, by keeping all these factors partially under control and by explicitly illustrating their partial differences, we provide sufficient basis to make the results comparable in a fair context.

3.2. The ReVolt (A conceptual autonomous ship) – System under analysis

DNV GL[2] is a leading classification society that is influential in shaping future maritime safety regulations and develops corresponding rules and standards for the classification of ships. It is crucial for DNV GL to comprehensively assess the potential safety and security risks of operating autonomous ships under diversely operational scenarios and unexpectedly environmental situations. DNV GL Group Technology and Research department has been developing a conceptual autonomous ship, called the ***ReVolt***, since 2014. The ReVolt has been dedicated to explore and experiment how to safely and securely integrate novel and emerging technologies/equipment into marine ships to achieve various degrees of automation.

In this paper, we select the ReVolt as the system under analysis to perform our comparative study. The ReVolt currently supports four operation modes, i.e., remote-controlled mode, dynamic positioning mode, emergency stop mode and autonomous navigation mode. In this paper, we limit our analysis scope to the ***autonomous navigation mode***. Under the autonomous navigation mode, the most critical accident that the ReVolt has to avoid is colliding with other objects, including vessels, swimmers, floating obstacles, or structures. To limit our analysis effort to a controllable level, we only focus on analysing the ***collision accidents under the autonomous navigation mode.*** Since the ReVolt is an unmanned ship and does not collect any confidential data during its missions, we do not consider security-related system-level losses due to confidentiality breaches in this study.

As shown in Fig. 7, various types of sensors and advanced functionalities have been integrated into the ReVolt to enable autonomous navigation. The notations used in Fig. 7 are explained in Table 2. Furthermore, the operational scenario under analysis and the possible collision accidents are described in Table 3, which specifies the scope of our comparative study performed by the two teams.

3.2.1. A brief overview of autonomous operation mode

Under the *autonomous operation mode*, the ReVolt will be able to receive a destination from the operator as input and move to the desired position autonomously. This will be accomplished by autonomously creating a safety path for the ReVolt to follow. When it navigates along the planned path, it will use cameras and a LiDAR to monitor its surroundings. These sensor measurements will be used for collision avoidance and to update the path plan.

The *navigation controller* is the most complex part of the ReVolt system and implements the major functionality of autonomous navigation. The navigation controller receives inputs from the operator via 4G communication network, the obstacle avoidance and the observer. The inputs include the desired destination, the position of any detected obstacles, and the current position, speed and heading of ReVolt. Then, the navigation controller provides a safety path as a set of waypoints to follow in order to safely reach the desired destination. Finally, it provides the force controller with the current position and orientation and the next desired position and orientation.

The *obstacle avoidance* receives the vision sensor data about the surroundings of the ReVolt. Those data are used to detect obstacles in the surrounding area and estimate the position, speed and heading of the detected obstacles. Those data are shared with the navigation controller.

The *observer* receives position and velocity sensor data to estimate the position, linear and angular velocity of the ReVolt. Those data are shared with the navigation controller. The ReVolt communicates with the onshore operator via the 4G communication network. It sends and receives messages in real-time. The operator mainly provides the ReVolt with its desired destination and control commands. The navigation controller uses the 4G network to send feedback information to the operator.

4. Analysis results

This section describes and compares the results obtained from the safety and security co-analysis of the ReVolt using the UFoI-E method and STPA-Extension.

4.1. UFoI-E results

4.1.1. Tailored CPS master diagram of the ReVolt

Fig. 8 illustrates the diagrammatic representation of the ReVolt as a tailored CPS master diagram. This tailored CPS master diagram is the result of a workshop adapting the generic CPS master diagram with the system specifications of the ReVolt. The team applying the UFoI-E method conducted this workshop using a shared widescreen and basic diagrammatic tools.

The tailored CPS master diagram provides a comprehensive model of the system that the team members used to communicate about the system and to analyse it systematically using CyPHASS. For the comparative study, this diagram emphasizes the components and interactions in the system relevant only for the autonomous mode of operation. Therefore, some parts of the master diagram are hidden with filters in the background and can be retrieved only if required.

At the physical layer (PL), the diagram shows the energy flow interactions between the thrusters and the vessel hull for the propulsion and steering of the vessel. The PL interacts with the physical environment, which poses energy disturbances (e.g. ocean currents, waves, and wind) as well as obstacles in the way (e.g. other vessels). At the cyber-physical layer (CPL), the diagram shows the information flows between the sensors, the computer on-board and the microcontrollers. The CPL implements the autonomous control actions that influence the PL in real-time feedback loops. At the cyber layer (CL), the diagram illustrates the remote workstation onshore, where the human supervisor issues commands via an HMI and uses a wireless network to communicate with the CPL, particularly with the computer onboard the vessel. Finally, the cyber and physical environments interact with the different layers of the system supplying information and energy flows, which may be subject to manipulations by the actions of hackers and saboteurs. These hackers and saboteurs could potentially penetrate the system layers to perform attacks. Similarly, malicious insiders could violate system guidelines and cause a sequence of UFoI-E.

4.1.2. Identification of scenarios and barriers in the CyPHASS bowtie

Fig. 9 summarizes the aggregation of scenarios identified with CyPHASS. To identify these scenarios, the team in charge conducted risk

[2] DNV GL is the independent expert in risk management and quality assurance: www.dnvgl.com

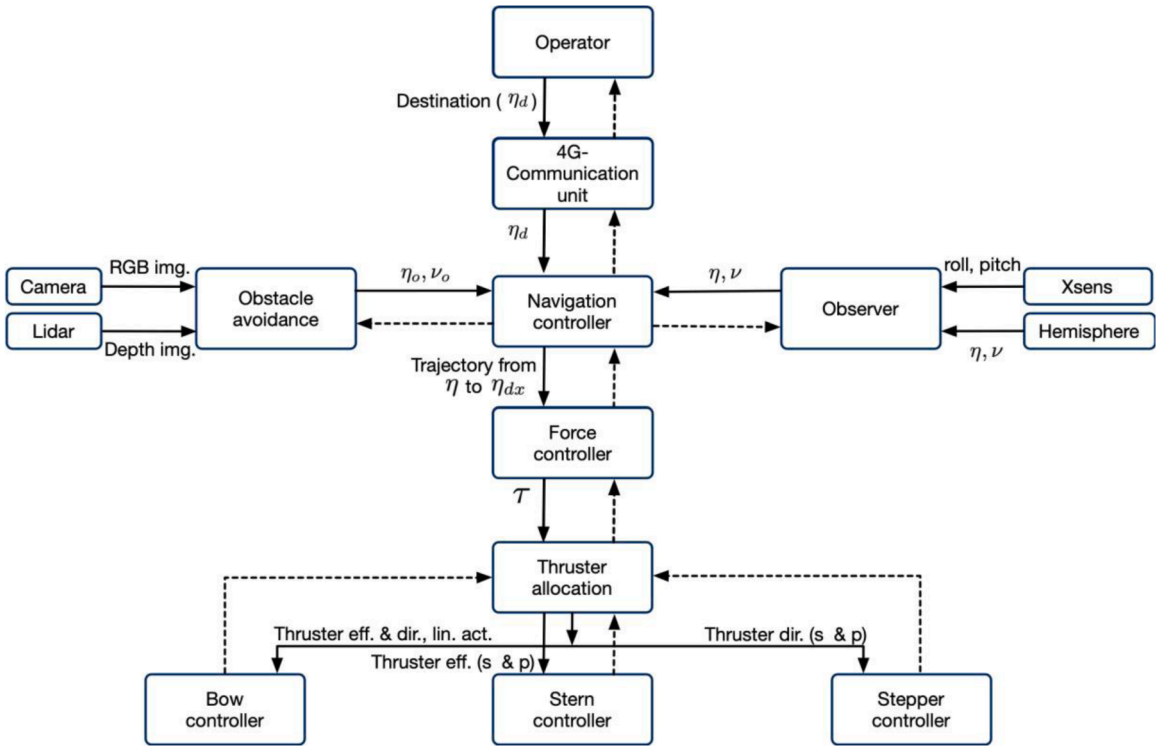

Fig. 7. Abstract control structure of the ReVolt. Courtesy of Solberg [41].

Table 2
Notations used in Fig. 7.

Notation	Explanation
η	Measured position and orientation (yaw)
η_d	The desired destination coordinate and heading
ν	Measured velocity and angular velocity
η_o	The estimated position and heading of a detected obstacle
ν_o	The estimated velocity of a detected obstacle
τ	Forces and moments vector
Trajectory from η to η_{d_x} :	The trajectory from the current position to the next point along the path
Thruster eff. (s & p)	Effort for the stern thrusters
Thruster dir. (s & p)	Direction for the stern thrusters

identification workshops using the CyPHASS database of checklists in spreadsheets format. The team applying the UFoI-E method conducted this workshop using a shared widescreen and basic spreadsheet software.

According to the scope of this comparative study, the scope of the analysis is a collision as an ultimate safety consequence. From this point, CyPHASS allows for systematic backward tracing of causes with their respective intermediate events according to the extended bowtie model. For each stage of the scenarios, the team also identified prevention and mitigation barriers according to the suggestions in CyPHASS. If the barriers were already present in the ReVolt, the team marked them as present (P). Conversely, the team marked the barriers as recommended (R) to add into the system.

For illustration purposes, Fig. 10 summarizes an extract of three scenarios in CyPHASS shown as three branches:

- Branch (1): extract of a scenario that initiates with a risk source at the physical layer.
- Branch (2): extract of a scenario that initiates with a risk source at the cyber-physical layer.
- Branch (3): extract of a scenario that initiates with a combination of risk sources at the cyber layer.

Fig. 10 shows how CyPHASS traces and links the different scenarios that converge into a unique consequence. Moreover, this figure demonstrates how barriers allocated at different stages of the scenarios act as layers of protection. If a risk source breaches a prevention barrier and leads to an intermediate event, the mitigation barriers to the right may mitigate the consequences of the intermediate event. Furthermore, if an

Table 3
The ReVolt's operational scenario under analysis.

ID	Operation	Accidents
1	Autonomous navigation mode	
1.a	Activation of autonomous mode	Another vessel collides with the ReVolt
1.b	Receive waypoint plan and destination from the human operator via an onshore control station	1) Another vessel collides with the ReVolt 2) The ReVolt drifts into other vessel or nearby structure
1.c	Follow the waypoint plan to the destination while avoiding collision with detected vessels/obstacles	1) Another vessel collides with the ReVolt 2) The ReVolt collides with another vessel 3) The ReVolt collides with the nearby structure
1.d	Reduce to zero speed at the destination	1) Another vessel collides with the ReVolt 2) The ReVolt drifts in other vessel or nearby structure 3)The ReVolt collides with nearby structure

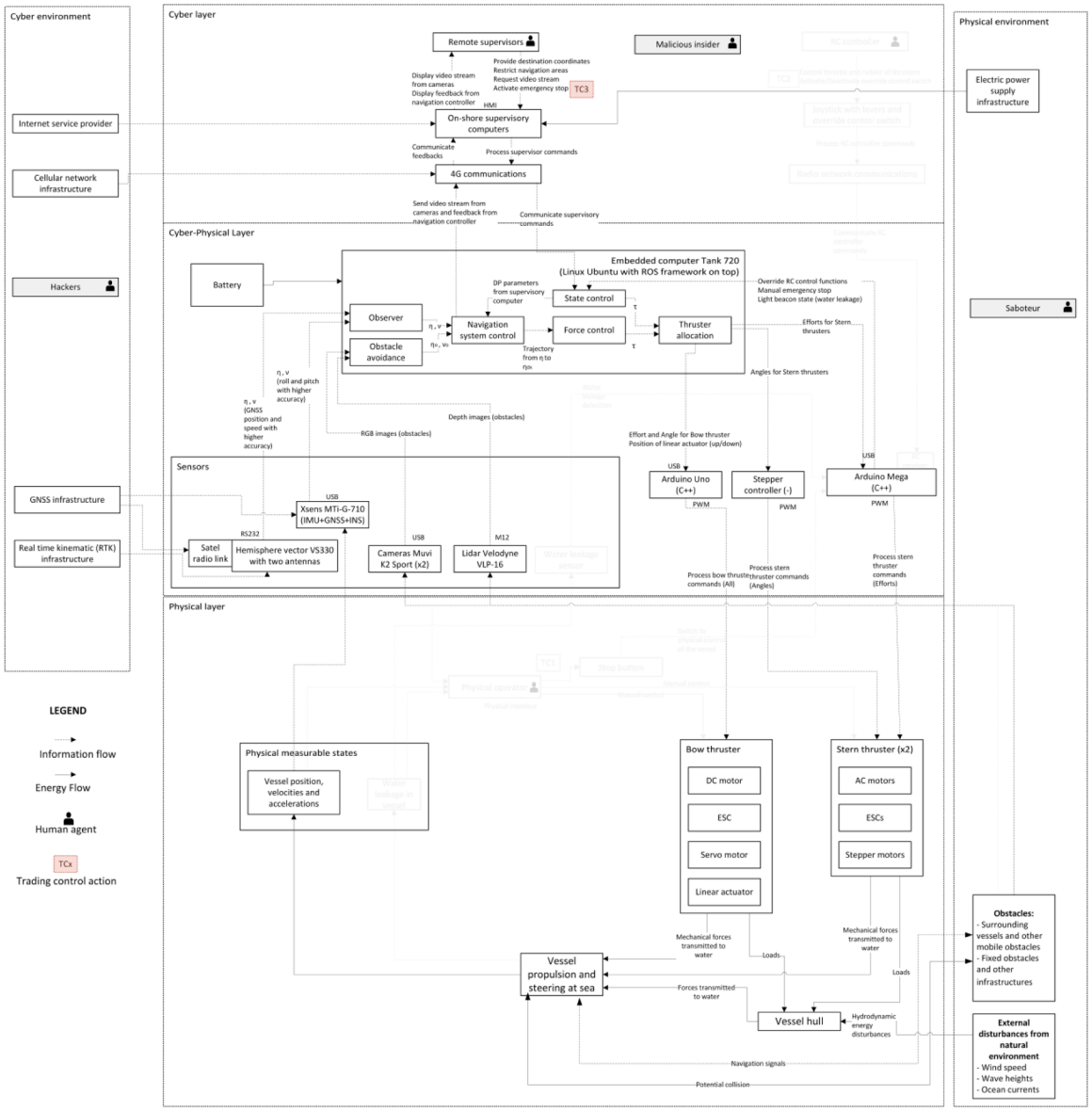

Fig. 8. Tailored CPS master diagram of the ReVolt (emphasis on autonomous mode of operation).

intermediate event breaches the mitigation barriers and causes a subsequent event, the mitigation barriers allocated at the subsequent stage of the scenario may still lead the system to a safe state. To simplify the visualisation of specific barriers, this figure only shows a single barrier at each stage of the bowtie.

In the complete set of scenarios in spreadsheet form, each element of a scenario has an assigned identifier (ID). These IDs trace the intermediate events identified in CyPHASS to their respective causes and consequences and their associated barriers.

4.1.3. Effort required in person-hours

As an indication of the effort required using the UFoI-E method to analyse the ReVolt system, Table 4 summarizes the person-hours that Team 1 spent in this analysis. Notice that, neglecting the preparation time reading the system specifications, the time designing the tailored CPS master diagram of the ReVolt was only seven person-hours. As expected, the most expensive part of the analysis was the identification of risk scenarios and the recommendation of barriers with CyPHASS. Still, the database of checklists in CyPHASS represented a benefit for Team 1 to complete this part of the analysis cost-effectively, not needing to spend time collecting documentation about lessons learned and checklists from external risk repositories. Moreover, the results of UFoI-E already include the recommendation of prevention and mitigation barriers acting as layers of protection. For a similar time of work and effort required, this recommendation of barriers is a benefit that is usually beyond the scope of STPA analyses.

4.2. STPA-Extension results

The beauty of performing an STPA analysis is to establish a traceability between a precedent (sub)-step and the following (sub)-steps by following a top-down approach. We demonstrate such traceability through presenting our analysis results by following the STPA-Extension workflow illustrated in Fig. 4. STPA-Extension is an iterative approach. All steps shown in Fig. 4 can be re-visited and the analysis results of each step can be revised accordingly.

4.2.1. Step 1: Define the purpose of this analysis

STPA-Extension starts from identifying system-level losses that must be prevented. Theoretically, conventional STPA and its variations can be

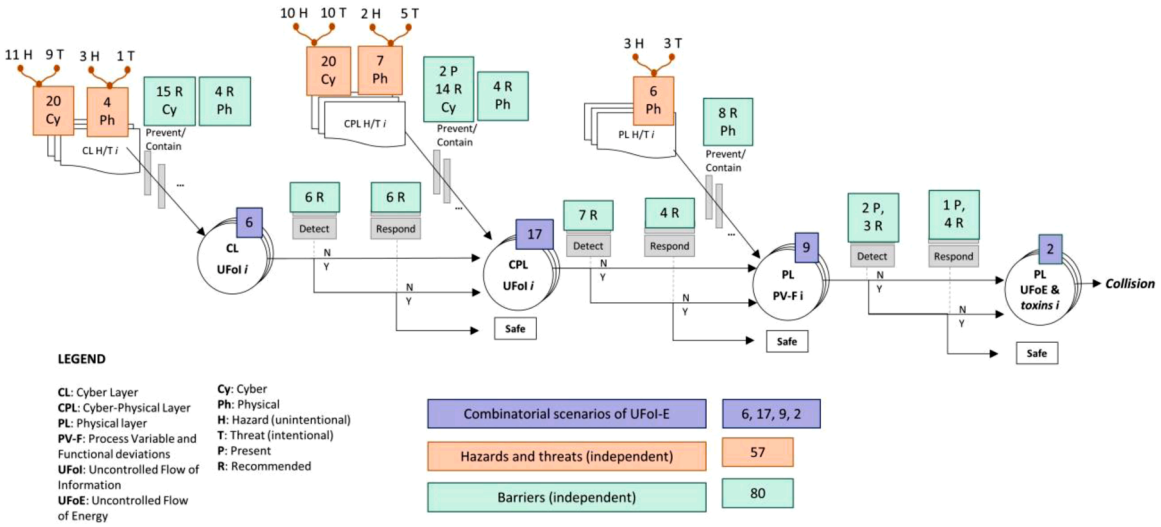

Fig. 9. Summary of aggregation of scenarios identified with CyPHASS.

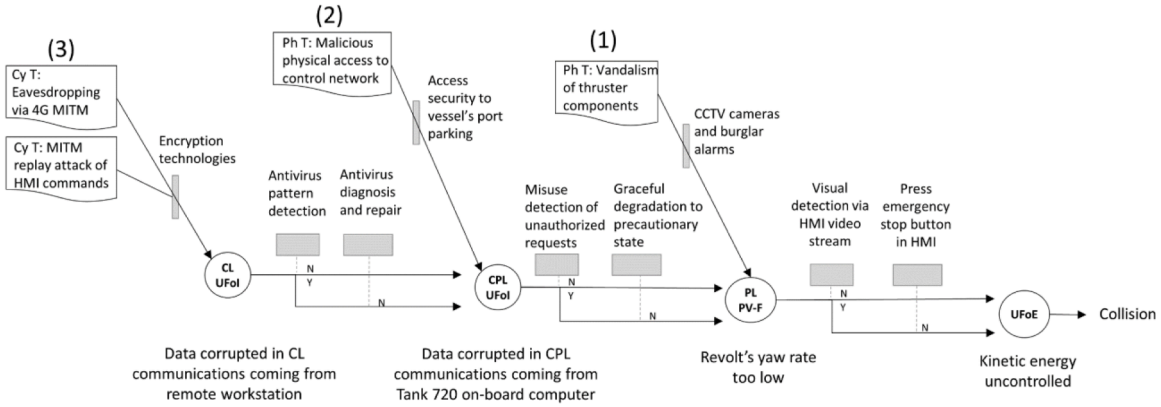

Fig. 10. Extract of three illustrative scenarios in CyPHASS, each with risk source origin at a different stage.

Table 4
Summary of person-hours used by Team 1 with the UFoI-E method.

Parts of the analysis	Time (person-hours)	%
Preparation reading system specifications	8	18%
System representation with tailored CPS master diagram	7	15%
Risk scenarios and barriers identification with CyPHASS	30.5	67%
TOTAL	**45.5**	**100%**

Table 5
Identified system-level losses[a].

Loss Index	Loss
L1	ReVolt efficiency loss
L2	Breach of COLREGs or too close to objects
L3	Damages to the ReVolt
L4	Loss of the ReVolt
L5	Damage to other vessels or structure
L6	Harm to people

[a] COLREGs: Convention on the International Regulations for Preventing Collisions at Sea, 1972.

applied broadly to not only safety losses but also losses related to security, privacy, performance and other system emergent properties. Practically, we have to define the system boundary and limit the scope of analysis to efficiently achieve desired results.

The operational scenario considered in this paper is to safely and securely navigate the ReVolt by following a predefined waypoint plan. Therefore, the scope of our analysis focuses on safety, security and performance-related losses. The identified system-level losses are listed in Table 5.

Losses are typically the result of accidents and often of a non-specific nature. As shown in Table 5, we may identify similarly generic losses for many other safety-critical systems as well. To bridge the gap between generic losses and further analysis, STPA-Extension re-introduces identification of system-level accidents, which are identified in a specific context or under specific operational scenarios [19]. Note that the system-level accident was originally included in the conventional STPA [31] but is replaced by "loss" in the latest version of STPA [32].

We chose one particular system-level accident to exemplify the complete analysis process of STPA-Extension due to time limitations when performing this comparative study. We further identified system-level hazards that could lead to this accident. Subsequently, we have to prevent the hazards from occurring or minimise the loss in case the hazards do occur by defining corresponding system-level constraints. The system-level accident, hazards and constraints identified in the study are listed in Table 6.

The last row in Table 6 establishes the traceability between one (sub)-step and the following (sub)-step(s). For example, one accident (i.e., A1) may lead to several losses (i.e., L3, L5 & L6). One hazard (i.e., H1)

Table 6
System-level accident, hazard & constraints of the ReVolt.

	Accident	Hazard	Constraint
Index	A1	H1	C1
Content	REVOLT turns/ drives/drifts into vessel/ object/structure	REVOLT navigates too close, or at too high speed towards other vessel, object or structure	REVOLT must keep a minimum distance and safe heading/ speed to other vessel, object or structure
Traceability	L3,L5,L6	A1	H1

may be related to one accident (i.e., A1). A system-level constraint is simply to invert a system-level hazard. Therefore, one constraint (i.e., C1) typically corresponds to one hazard (H1).

Step 1 of the conventional STPA stops here. However, STPA-Extension introduces an additional sub-step, i.e., functional requirements, to facilitate modelling the control structure of the ReVolt. As shown in Table 7, we finished Step 1 analysis by specifying a list of functional requirements to fulfil constraint C1.

4.2.2. Step 2: Model control structure of the ReVolt

A hierarchical control structure is a system model that is composed of feedback control loops [32]. The control structure consists of two important control system properties, i.e., controllability (control input and control actions) and observability (feedback and process model) [19].

We began with an abstract control structure (refer Fig. 7) and iteratively refined it. More specifically, we identified controllers, the corresponding control actions of each controller and the feedback of each control action by analysing functional requirements specified in Table 7. Those artefacts identified during control structure modelling are summarized in Table 8.

4.2.3. Step 3: Identify unsafe control actions (UCAs)

Identification of unsafe control actions (UCAs) plays a vital role in STPA analysis. The definition of an UCA is a control action that, in a particular context and worst-case environment, will lead to a hazard [32]. To consistently identify UCAs for different systems, four types of UCAs are defined in the conventional STPA. Moreover, we introduced two additional types of UCAs to properly describe the cases in our study. Particularly, UCA3 is implicitly covered in UCA2, but we introduce it in STPA-Extension to explicitly represent this type of UCA. The complete list of UCAs adopted in our study is described in Table 9, where UCA3 and UCA6 are the two additional types introduced in this study.

The practice of identifying UCAs is to enumerate all possible combinations of the controller, control action, and UCA type, which may lead to a hazard. Table 8 lists three controllers and five control actions. Note that UCA5 does not apply to our study and was excluded from UCA identification. The possible UCAs identified by combining the controller and control action is summarized in Table 10.

Table 7
Functional requirements derived from constraint C1.

Index	Functional Requirement	Traceability
F1.1	REVOLT must know its safe minimum distance to other vessels, objects or structures	C1
F1.2	REVOLT must know the position, speed and heading of other vessels	C1
F1.3	REVOLT must predict the future path/position, speed and heading of other vessels	C1
F1.4	REVOLT must decide its safe heading and speed to keep the safe minimum distance to other vessels, objects or structures	C1
F1.5	REVOLT must control its heading and speed within the safe heading and speed settings	C1
F1.6	REVOLT must know its position, speed and heading	C1

4.2.4. Step 4: Identify causal scenarios

A causal scenario describes the causal factors that can lead to the unsafe control actions and hazards [32]. To explicitly identify initial causal factors for each unsafe control action identified in Step 3, we conducted the causal scenario identification in a step-wise fashion as illustrated in Fig. 4.

Firstly, we started with examining each UCA listed in Table 10 and enumerated the potential effects of all possible failures that could result in this UCA by following four general types of Accident Causes (ACs) listed below.

- AC1 – Wrong or missing control input
- AC2 – Wrong control logic
- AC3 – Wrong or missing situation understanding, feedback/sensing
- AC4 – Lost/altered control action or actuation

Note that the four types of ACs are adapted from Leveson and Thomas [32] but customized to more precisely capture hazards and threats of the system under analysis. Remarkably, the sole combination of UCA6 and AC4 is reasonable.

One causal scenario was specified to describe one specific effect and the resulted UCA. One example of causal scenarios is:

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

Secondly, we listed all possible failure modes for the specific effect described in each causal scenario and named the failure modes as ***Design-specific causes (DSC)*** in this paper. Following the causal scenario example given above, we identified five DSCs for **CS-1** as listed in Table 11.

Lastly, for each DSC listed in Table 11, we enumerated all initial causal factors, which could contribute to the occurrence of this DSC by following three generic cause categories: ***accidental, unintentional, and intentional***. Both accidental and unintentional causes belong to safety category, while intentional causes are of security category. The accidental causes are inherent in the design of an element (e.g. component or functionality), such as design defects, software implementation errors, hardware failures, communication failures, etc. The unintentional causes are generally due to human errors. For instance, the human operator forgets to send a control command to the system under control or unintentionally configures a wrong value of some parameters. The intentional causes represent all possibilities that a malicious entity could deliberately compromise a specific element by exploiting the security vulnerabilities of this element.

The summary of causal scenarios analysis for hazard H1 is given in Table 12. One casual scenario may correspond to more than one design-specific causes. In addition, one DSC may be traced to several initial causal factors that are still at a very generic level. We can apply traditional safety or security analysis methods to model concrete root causes for each (generic) initial factor. However, we did not reach this level of details in our study due to time limitations.

4.2.5. Effort spent on STPA-Extension analysis

The STPA-Extension analysis team divided the analysis work into two parts according to the competence of the two analysts. The first analyst, who has extensive knowledge of cybernetics and two years of experience in applying STPA to safety analysis, was mainly responsible for the tasks from Step 1 to design-specific causes in Step 4. The second analyst, who had a certain knowledge of the ReVolt and two years of experience in safety and security co-analysis, mainly worked on categorising initial causal factors into three generic groups and applying such categorisation to initial causal factor analysis of each design-specific cause. The total effort spent on the STPA-Extension analysis is summarized in Table 13.

Table 8
Control structure analysis results.

ID – Controller name	Responsibility of controller	ID – Control actions provided	To sub-controller or actuator	Feedback
C1 – Operator	Command waypoint plan to ReVolt	CA1 – Waypoint plan	Navigation controller	NA
C1 – Operator	Command safe collision avoidance distance to ReVolt	CA2 – Safe distance	Navigation controller	NA
C2 – Navigation controller	Collision avoidance control	CA1 – deviating waypoint plan	Navigation Controller – Trajectory control	Other vessel's position/speed/ heading Other vessel's future position/speed/ heading ReVolt's own position/speed/ heading
C2 – Navigation controller	Trajectory control	CA2- trajectory	Force control	Applied force and direction
C3 – Force controller	Calculate needed thrust force	CA1- thrust force/ momentum	Thruster allocation	Applied thrust force/momentum

Table 9
Six types of Unsafe Control Actions.

ID	Unsafe control action (UCA)
UCA1	Not providing the control action leads to a hazard
UCA2	Providing the control action leads to a hazard
*UCA3	*Providing a potentially safe control action but with too high/low value*
UCA4	Providing a potentially safe control action but too early/late, or in the wrong order
UCA5	The control action lasts too long or is stopped too soon (only for continuous control actions)
*UCA6	*Providing the control action but it is not followed by the receiver*

Table 10
Possible UCAs identified.

Combination of Controller & Control action	Possible UCAs
C1 - Operator & CA1- Waypoint plan	UCA1, UCA2, UCA6
C1 - Operator & CA2 – Safe distance	UCA1, UCA2, UCA3, UCA6
C2 - Navigation controller & CA1 – Deviating waypoint plan	UCA1, UCA2, UCA3, UCA4, UCA6
C2 - Navigation controller & CA2 – Trajectory	UCA1, UCA2, UCA4, UCA6
C3 – Force controller & CA1 – Thrust force/ momentum	UCA1, UCA2, UCA3, UCA6

4.3. Comparison of results

Based on the results previously described for each safety and security co-analysis method, we conducted a comparison of results in terms of relative completeness obtained by each method. Table 14 summarizes the relative completeness from the STPA-Extension and the UFoI-E analyses of the ReVolt. This table counts the number of risk scenarios and classifies them according to the relative completeness framework described in Section 3. Similarly, Fig. 11 illustrates the results in our Venn diagram framework.

To better illustrate these results, Table 15 shows a representative extract of the comparison. This table describes specific risk scenarios identified only by one of the methods and specific risk scenarios identified by both methods. Note that this table shows at least one scenario for each of the cases previously described in the Venn diagram. Moreover, for each scenario, there is an explanation of the reason for diverging results or the common factors in the correspondence of results. Notice that the STPA-Extension method does not suggest barriers to prevent and mitigate the occurrence of the scenarios, being beyond the scope of the analysis. Therefore, we did not include in the comparison the prevention, detection and response barriers identified in UFoI-E. For more details into the criteria for comparison, see the discussions in Section 5.2.

From this comparison of results, we can argue that, for a similar degree of effort required to perform the analyses, both UFoI-E and STPA-Extension achieved a similar level of completeness in the safety and security co-analysis of the ReVolt. Nevertheless, this similar completeness in terms of the number of scenarios identified is partially different in terms of the scope of the results obtained by each method. Based on these results, the following section discusses our propositional generalisations about the strengths and limitations of the methods used.

Table 11
Design-specific causes and initial causal factors for Causal scenario - 1.

DSC Index	DSC description	Causal factors (A: accidental; I: Intentional)
DSC-1	Object detection model (i.e., a convolutional neural network) does not detect other vessel/ obstacle.	A1 - Object detection model fails to detect vessel/object due to its error-prone 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.
DSC-2	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.
DSC-3	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.
DSC-4	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 12
Summary of causal scenarios analysis.

Hazard (ID: H1)	Number of identified causal scenarios	Number of Design-specific causes (DSC)	Number of initial causal factors (generic)
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**

Table 13
STPA-Extension analysis effort statistics.

Parts of the analysis	Time (person-hours)	%
Preparation reading system specifications	12	23.5%
Step 1 – Step 3	12	23.5%
Step 4 (Causal scenario, Design-specific cause, Initial causal factors)	27	53%
TOTAL	**51**	**100%**

Table 14
Summary of relative completeness results.

Total risk scenarios	115	100%
STPA-unique results	38	33%
Team specific STPA-unique results	9	8%
Method-specific STPA-unique results	29	25%
UFoI-E-unique results	36	31%
Team specific UFoI-E-unique results	6	5%
Method-specific UFoI-E-unique results	30	26%
Correspondence in results	41	36%

5. Discussion

5.1. Threats to the validity of this case study

The results from this case study rely on the experience of two risk analysis teams performing safety and security co-analysis of a single common system. Therefore, the aim of this case study is not statistical validity. Instead, this case study research allows for the discovery of propositional generalisations [42]. Using inductive reasoning, from this case study we can derive propositions that are relevant beyond our specific domain of application [18]. Namely, we argue that these generic propositions apply for other risk analyst groups using the UFoI-E method and STPA-Extension to analyse other systems.

For the relative evaluation of completeness, we separate the reasons leading to diverging results in our analyses in two classes of reasons:

1 Team-specific: reasons associated with the team deploying the method
2 Method-specific: generic reasons associated with the safety and security co-analysis methods as such.

We argue that this separation is necessary to account for two main team-specific factors.

First, the execution of the method may lead to incomplete results because the team may have carried out the analysis imperfectly. This imperfect execution can lead to oversights in risk identification. The causes of imperfect execution can be a lack of knowledge of the system under analysis, or because the method was not perfectly applied by the team in some specific sections [44]. To partially control against these analysis imperfections, the two teams provided their level of knowledge of the ReVolt and their years of experience as safety and security analysts. It is worth to emphasize that STPA-Extension can be applied to a very broad view of accident mechanism concerning social, environmental, human, technological, and contextual factors. To leverage the analysis scope to an appropriate level where both UFoI-E and STPA-Extension could generate comparable results, the STPA-Extension team had to narrow the system boundary and excluded non-technical aspects from the analysis.

Second, the execution of the method may also lead to incomplete results because one team dedicated more resources and effort to perform the analysis. Namely, software tools, time of work and previous analyses available can influence the relative level of completeness achieved by a team using a method. In addition, the analysts may subjectively make a trade-off to reach a certain level of analysis granularity under time limitations. This is a common practice in industry when conducting a risk analysis for a complex system. To partially control against resources and effort required, each team recorded the number of hours spent performing each phase of the risk analysis. We conceived the number of hours used as one indicator of the effort required. Furthermore, the teams reported the software tools used to perform their analyses, which were limited to typical software drawing and spreadsheet tools without automation features. Since the team members are knowledgeable in the use of their particular methods, this case study does not discuss the effort

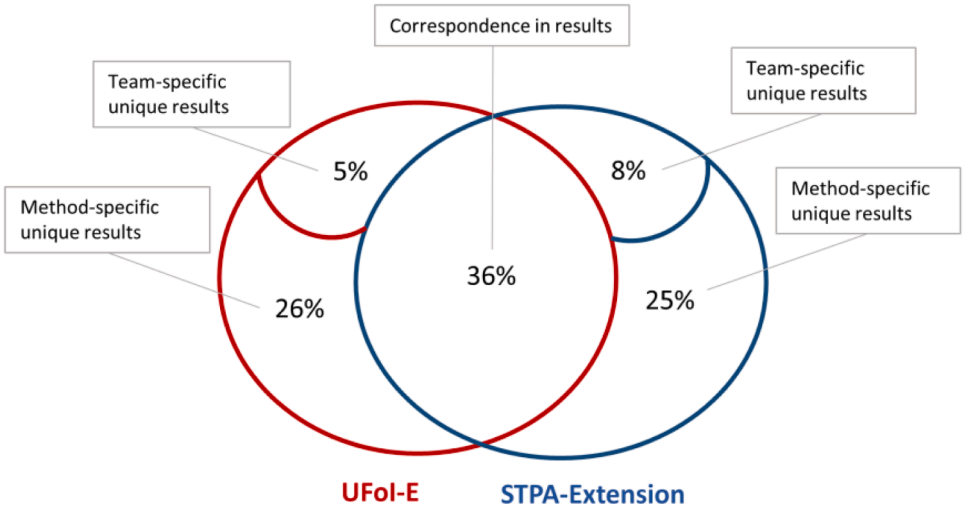

Fig. 11. Relative completeness results in Venn diagram framework.

Table 15
Illustrative examples comparing the results from the STPA-Extension and the UFoI-E analyses.

ID	STPA Causal scenario	UFoI-E harm scenarios (without barriers)	Comment
28	**Navigation** (Controller 2) **does not calculate deviating waypoint plan** (Unsafe control action 1) because **ReVolt's collision avoidance algorithm predicts own future position too far away from other vessel/ object** (Design-specific case). This leads to that **ReVolt navigates too close, or at too high speed towards other vessel, object or structure** (hazard: H1).	NOT FOUND	**STPA-unique result** – ***Team-specific reason:*** *Team's in-depth* ***knowledge*** *of collision avoidance in the ReVolt*
34	**Navigation** (Controller 2) **provides too little/ large deviating waypoint plan** (Unsafe control action 3) because **obstacle avoidance functionality estimates other vessel's heading wrongly to the "safe side" or "unsafe side"** (Design-specific cause). This leads to that **Revolt navigates too close, or at too high speed towards other vessel, object or structure** (Hazard 1).	NOT FOUND	**STPA-unique result** – ***Method-specific reason:*** *STPA's special attention to the control algorithm and its* ***process model*** *leading to unsafe control actions as a control problem*
104	NOT FOUND	**Saboteur intentionally misaligns or moves physical components and connections of vessel to alter mechanism behaviour** (PL threat 2) → **Vessel's linear speed remains too high when required to decrease** (PL F deviation 1) → **Vessel's kinetic energy uncontrolled leading to collision** (UFoE 1)	**UFoI-E-unique result:** – ***Team-specific reason:*** *Level of analysis in UFoI-E covers* ***physical security*** *threats, not covered in the STPA team's scope of analysis*
129	NOT FOUND	**Blackout leading to power supply loss in on-shore supervisory station** (CL hazard 22) → **Trading control decision (emergency stop) from on-shore human supervisor delayed or unavailable** (CL UFoI 6) → **Control communications of emergency stop delayed or unavailable coming from Tank 720 computer** (CPL UFoI 11) → **Vessel's linear speed decreases too late** (PL F deviation 4) → **Vessel's kinetic energy uncontrolled leading to collision** (UFoE 1)	**UFoI-E-unique result:** – ***Method-specific reason:*** UFoI-E's special attention to **resources** provided from the physical and cyber environments as flows of information or energy (e.g. power supply)
2	**Human operator** (Controller 1) **provides unsafe waypoint plan** (Unsafe control action 2) because **operator has wrong situational understanding** (Design-specific cause). This leads to that **ReVolt navigates too close, or at too high speed towards other vessel, object or structure** (hazard: H1).	**Human-machine interfaces corrupted (e.g. display incorrect information, withhold the display of correct information; see Maroochy Water breach, Stuxnet)** (CL UFoI 2) → **Data corrupted in Tank 720 computer** (CPL UFoI 1) → **Vessel's linear speed remains too high when required to decrease** (PL F deviation 1) → **Vessel's kinetic energy uncontrolled leading to collision** (UFoE 1)	**Correspondence of results:** – *Example of risk associated with on-shore human supervisor*
18	**Navigation** (Controller 2) **does not provide waypoint plan** (Unsafe control action 1) because **LIDAR range is offset further than actual** (Design-specific cause). This leads to that **Revolt navigates too close, or at too high speed towards other vessel, object or structure** (Hazard 1).	**Sensors data from physical layer or environment corrupted** (CPL UFoI 6) → **Vessel's linear speed remains too high when required to decrease** (PL F deviation 1) → **Vessel's kinetic energy uncontrolled leading to collision** (UFoE 1)	**Correspondence of results:** – *Example of risk associated with sensors*
29	**Navigation** (Controller 2**) provides unsafe deviating waypoint plan** (Unsafe control action 2) because **safe distance parameter is reduced over communication** (Design-specific cause). This leads to that **Revolt navigates too close, or at too high speed towards other vessel, object or structure** (Hazard 1).	**Man-in-the-middle replays of feedback data or HMI control commands** (CL threat 9) → **Data corrupted during cyber communications in cyber layer** (CL UFoI 1) → **Data corrupted in computer Tank 720** (CPL UFoI 1) → **Yaw rotation speed remains too low (No function)** (PL F deviation 6) → **Vessel's kinetic energy uncontrolled leading to collision** (UFoE 1)	**Correspondence of results:** – *Example of risk associated with wireless communications*

required to familiarise a new team with the terminologies and tools of each method.

5.2. Ensuring mutual correspondence between the results in UFoI-E and STPA-Extension

In our experience, it was not straightforward to decide beforehand

how to compare the analysis results of UFoI-E and STPA-Extension regarding the evaluation criterion of completeness. The reasons include the terminology discrepancy between the two methods, the different execution stages defined in each method, and the different supporting tools used by each method. Although we conducted several workshops to discuss the differences in terminologies, we could not agree on the theoretical correspondence between the results obtained from the two methods. Therefore, we agreed to perform the analyses as each method recommends, i.e. without a predefined template to compare the results. After we obtained the results from each method, we were able to identify more clearly the correspondence between the results obtained in each method.

Table 16 illustrates the approximate correspondence between the terminologies used in each method. As a main difference, note the definition of the term "hazard". In STPA-Extension, based on Leveson [30], a hazard is an unsafe system state leading to a loss. In UFoI-E, based on the Society for Risk Analysis Glossary [1,2], a hazard is an unintentional or natural risk source, in opposition to a threat that is an intentional risk source. Therefore, a comparison of the "hazards" identified in each method would not have been a suitable comparison. Instead, our comparison was based on the common meaning of the terms that have a direct correspondence between the two methods.

Furthermore, in Table 17 we describe the correspondence between the stages of the analysis in each method. Namely, this table compares the stages that show how each method identifies the risk scenarios. As the scope of STPA-Extension does not include the recommendation of barriers, we did not include in the comparison the prevention, detection and response barriers documented in the UFoI-E results. Finally, this table summarizes our suggestions to combine each method with some corresponding features of the other method, exploiting the strengths of each method and allowing for a more comprehensive risk identification result.

5.3. Suggestions to combine features of UFoI-E and STPA-Extension

From the system representation and modelling perspective, STPA-Extension is an iterative method that begins from a simple system-level control structure and continue revising the control structure until reaching the desired level of system abstraction. This simple control structure representation applies for many types of engineered systems and it is not exclusive for CPSs. The UFoI-E method, instead, provides the CPS master diagram as a predefined template to represent the specific architecture of the CPS under analysis. Therefore, even if both STPA-Extension and the UFoI-E method can be applied to analyse CPSs, the UFoI-E method is specifically tailored for this class of systems and STPA-Extension is more generic.

STPA-Extension is a top-down risk analysis approach for a broad range of socio-technical systems as well as CPSs. It can guide analysts to perform an initial round analysis without acquiring too much prior domain or system-specific knowledge. It allows iteratively adding more details into each step to make the analysis more complete as the analysts gain more knowledge or acquire more information about the system under analysis. Moreover, causal scenarios capture the dynamism of the system under analysis by combining the worst-case environmental conditions, system state, failure modes, and the interactions among all involved components/functions along the control path.

Table 16
Correspondence between the terminologies in UFoI-E and STPA-Extension.

Common meaning of the term	Safety and security co-analysis method	
	UFoI-E	STPA-Extension
Unsafe system state, leading to loss	Uncontrolled Flows of Energy (UFoE)	Hazards
Deviations, leading to unsafe system state	a) Process Variable and Functional deviations (PV-F) b) Uncontrolled Flows of Information (UFoI)	a) Unsafe control actions (UCAs) b) Design-specific causes (DSCs)
Risk sources, leading to deviations	a) Hazards (unintentional risk sources) b) Threats (intentional risk sources)	Causal factors

The comparative study reveals that STPA-Extension can be enhanced by combining the strengths of UFoI-E or other safety/security analysis methods. More specifically, we propose three possible improvements listed below.

1 A generic list of sources of hazards and types of accidents can be helpful in STPA step 1 to find a more complete list of accidents and STPA hazards.
2 The CPS master diagram of UFoI-E can be beneficial in abstracting the control structure efficiently. However, to avoid adding too many details too early, the analyst has to carefully decide the right abstraction level for which the CPS diagram is suitable.
3 As we have introduced the generic categories of initial causal factors, the checklist of hazards and threats developed in UFoI-E can be referred to concretize the specific initial causal factors for each element. Moreover, we can combine other suitable safety/security analysis methods to better cover and investigate relevant causes for the unsafe control scenarios.

Similarly, the UFoI-E method can benefit from the results obtained using STPA-Extension in two generic ways:

1 In parallel with the system conceptualisation phase with the CPS master diagram, the analysts could develop a detailed explanation of the context of losses as in STPA step 1. In terms of the CPS master diagram, this context of losses would be a description of the different interactions between the physical layer of the CPS and the physical environment in different circumstances.
2 The CyPHASS database has the capacity to learn with new inputs and expert feedbacks. Therefore, from the comparison of results obtained in this case study, the CyPHASS database was enhanced with an expanded checklist of CPL UFoI. We identified some cases from the STPA results that were not suggested to consider in the CyPHASS database. Particularly, new considerations for autonomous navigation control algorithms and machine learning technologies were included as CPL UFoI, together with their generic risk sources and suggested barriers.

In summary, both STPA-Extension and the UFoI-E method proved to be suitable risk identification methods for safety and security co-analysis. Each method has some particular features that facilitate a similar level of completeness. However, to obtain a higher and more reliable level of completeness, we suggest that risk analysts can benefit from the combined results of applying the two methods in their studies. Particularly for teams of experts specialised in one of these methods, we provide tailored recommendations to combine specific parts of the other method to enhance and validate their results. These tailored recommendations support a more complete analysis with a cost-efficient approach.

6. Conclusions

In this paper, we objectively assessed the feasibility and efficiency of applying the UFoI-E method and STPA-Extension to safety and security co-analysis through an empirical study carried out by two independent analyst teams. We purposely defined the scope of a common system under analysis (i.e., the ReVolt, a conceptual autonomous ship) to leverage the advantages of the two methods.

Furthermore, we carefully defined evaluation criteria to formulate an original comparative framework where the results obtained from the two methods can be interpreted properly. This comparative framework

Table 17
The stages of each method, their mutual correspondence, and potential for combination.

Phase of the analysis	Parameters	Risk identification method UFoI-E	STPA-Extension	Comparison	Potential for combination
(1) Background for the analysis	(1.1) Context of losses	Potential cases of UFoE identified at the physical layer and the physical environment of the CPS master diagram	Identification of consequences, accidents, hazards/constraints, and functional requirements	- STPA-Extension considers a wider range of consequences (not only safety-related) and guides the analysts to systematically identify them	- In UFoI-E, include a detailed explanation of the context when defining the CPS master diagram
	(1.2) System representation and diagrams	Detailed system representation of the CPS and its environments using the CPS master diagram as a framework	Model the generic control structure of the system and refine it from analysis iterations	- UFoI-E provides more level of detail to assist the analysts in drawing a comprehensive control structure of the CPS. - In STPA-Extension, the system under analysis could be different from a CPS, so it is more generic in scope.	- In STPA-Extension, use the CPS master diagram to review the control structure of the CPS and refine it. Still, one can have the risk of adding too much detail too early.
(2) Analysis	(2.1) Identification of unsafe system state	Using the CyPHASS database: Selected for the specific system using the database of uncontrolled flows of energy (UFoE)	Defined in the context of losses (1.1)	- In UFoI-E, the method assists the analysts to check how different energy sources in the system could lead to physical harm - In STPA-Extension, the sources of harm are inferred from the system level losses in a systematic way	In STPA-Extension, a generic list of sources of hazards and types of accidents can help to find a more complete list of accidents and STPA hazards
	(2.2) Identification of deviations leading to unsafe system states	Using the CyPHASS database: Selected for the specific system using the database of process variable and functional deviations (PV-F) at the physical layer, and from the uncontrolled flows of information (UFoI) at the cyber and cyber-physical layers.	Using STPA deviation guidewords: Inferred from inverting the control requirements into unsafe control actions (UCAs)	- In UFoI-E, the conditions leading to harm are not uniquely associated with the actions of a controller. They can also be the result of sequential deviations in the performance of specific components of the system. - In STPA-Extension, the analyst applies the deviation guidewords to each control action as a hierarchic process. Arguably, STPA-Extension assists the analysts in a deeper study of the control algorithm and process model of the controllers.	- In UFoI-E, enhance the CyPHASS database with an expanded checklist of CPL UFoI related to process models. The CyPHASS database has the capacity to learn from new inputs and expert feedbacks.
	(2.3) Identification of risk sources leading to deviations	Using the CyPHASS database: - Selected for the specific system using the databases of risk sources (hazards and threats) at each layer of the system	Identification of causal factors at different sections of the control structure	- In UFoI-E, the analyst can use the extensive checklists of hazards and threats derived from lessons learned and apply it to their specific CPS architecture. - In STPA-Extension, the causal factors reach the level of generic risk sources, without further detail into underlying causes. For in-depth causal analysis, the analysts need to refer to other security analysis methods, e.g. attack trees.	- In STPA-Extension, use the CyPHASS database to complement the causal analysis of both unintentional and intentional causal factors, including explicit reliance on resources.
	(2.4) Recommendations for system protection and countermeasures	Using the CyPHASS database: Selected for the specific system using the databases of prevention barriers against threats/hazards at each layer, as well as detection and response barriers for each deviation leading to harm.	Design decisions to prevent UCAs in the control structure (2.2).	- In UFoI-E, there is a comprehensive set of countermeasures to protect the system, using both prevention and mitigation strategies in the bowtie convention.	N/A

facilitates the replicability of comparative studies assessing the capabilities of safety and security co-analysis methods. We selected two fundamental aspects, which are *completeness* and *analysis effort,* to evaluate the efficiency of two methods. More specifically, we used the *relative level of completeness* to assess the results obtained from this comparative study because no benchmark results of the ReVolt is available for calculating the absolute level of completeness.

Correlating the completeness of analysis results with analysis effort demonstrates that the two methods yield similar efficiencies. However, we must be aware that there is a significant discrepancy of risk scenarios identified by the two methods. This is mainly caused by two reasons: (1) the unique strengths and weaknesses of each method, and (2) the variety of each analyst team's domain knowledge and analytical skills. We further mutually mapped the results of one method to those of the other. Such mapping reveals insights into the derivation of propositional generalisations and discussion of the strengths and limitations of UFoI-E and STPA-Extension.

In the end, we propose a more comprehensive co-analysis framework to combine the UFoI-E method and STPA-Extension, which exploits their unique strengths and allow for a more complete and cost-effective safety and security co-analysis of cyber-physical systems.

CRediT authorship contribution statement

Nelson H. Carreras Guzman: Conceptualization, Methodology, Validation, Investigation, Data curation, Writing - original draft, Writing

- review & editing, Supervision. **Jin Zhang:** Validation, Investigation, Writing - review & editing. **Jing Xie:** Conceptualization, Validation, Investigation, Writing - original draft, Writing - review & editing, Supervision. **Jon Arne Glomsrud:** Validation, Investigation, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Amundrud Ø, Aven T, Flage R. How the definition of security risk can be made compatible with safety definitions. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2017;231(3):286–94. https://doi.org/10.1177/1748006x17699145.

[2] Aven T, Ben-Haim Y, Andersen HB, Cox T, Lopez Droguett E, Greenberg M, Zio E. Society for Risk Analysis Glossary 2018. https://doi.org/10.4135/9781412959216.n276.

[3] Body OHS, Models K, April S. *Models of Causation: Safety. OHS Body of Knowledge*. Safety Institute of Australia Ltd; 2012. Retrieved from, http://www.ohsbok.org.au/wp-content/uploads/2013/12/32-Models-of-causation-Safety.pdf.

[4] Carreras Guzman, N. H. (2020). CyPHASS prototype: Cyber-Physical Harm Analysis for Safety and Security. Retrieved November 11, 2020, from https://orbit.dtu.dk/en/projects/cyphass-prototype-cyber-physical-harm-analysis-for-safety-and-sec.

[5] Carreras Guzman NH, Kozine I. Uncontrolled flows of information and energy in cyber-physical systems. European Safety and Reliability Association Newsletter 2018:2–3. SeptemberRetrieved from, http://www.esrahomepage.eu/filehandler.ashx?file=16438.

[6] Carreras Guzman NH, Kozine I, Lundteigen MA. An integrated safety and security analysis for cyber-physical harm scenarios (Manuscript under review). Safety Science 2020.

[7] Carreras Guzman NH, Kufoalor DKM, Kozine I, Lundteigen MA. Combined safety and security risk analysis using the UFoI-E method: A case study of an autonomous surface vessel. In: Proceedings of the 29th European Safety and Reliability Conference (ESREL 2019); 2019. p. 4099–106. https://doi.org/10.3850/978-981-11-2724-3_0208-cd.

[8] Carreras Guzman, N. H., & Mezovari, A. G. (2019). Design of IoT-based Cyber-Physical Systems: A Driverless Bulldozer Prototype. *Information*, 10(11). https://doi.org/10.3390/info10110343.

[9] Carreras Guzman NH, Wied M, Kozine I, Lundteigen MA. Conceptualizing the key features of cyber-physical systems in a multi-layered representation for safety and security analysis. Systems Engineering 2020;(23):189–210. https://doi.org/10.1002/sys.21509.

[10] Chockalingam S, Hadziosmanovic D, Pieters W, Teixeira A, van Gelder P. Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications. In: Critical Information Infrastructures Security. 8th International Workshop, CRITIS 2013. Revised Selected Papers: LNCS 8328. Vol. 8328; 2013. https://doi.org/10.1007/978-3-319-03964-0.

[11] Creswell JW. Research design: Qualitative, quantitative, and mixed methods approaches. 4th ed. Thousand Oaks, CA, USA: Sage; 2014.

[12] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly: Management Information Systems 1989;13(3):319–39. https://doi.org/10.2307/249008.

[13] de Ruijter A, Guldenmund F. The bowtie method: A review. Safety Science 2016;88:211–8. https://doi.org/10.1016/j.ssci.2016.03.001.

[14] Dunjó J, Fthenakis V, Vílchez JA, Arnaldos J. Hazard and operability (HAZOP) analysis. A literature review. Journal of Hazardous Materials 2010;173(1–3):19–32. https://doi.org/10.1016/j.jhazmat.2009.08.076.

[15] Ericson CA. Hazard Analysis Techniques for System Safety. John Wiley and Sons, Inc; 2005. https://doi.org/10.1002/0471739421.

[16] Friedberg I, McLaughlin K, Smith P, Laverty D, Sezer S. STPA-SafeSec: Safety and security analysis for cyber-physical systems. Journal of Information Security and Applications 2017;34:183–96. https://doi.org/10.1016/j.jisa.2016.05.008.

[17] Gibson JJ. The contribution of experimental psychology to the formulation of the problem of safety – a brief for basic research. In: Behavioral Approaches to Accident Research. Association for the Aid of Crippled Children; 1961. p. 77–89.

[18] Gioia DA, Corley KG, Hamilton AL. Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. Organizational Research Methods 2013;16(1):15–31. https://doi.org/10.1177/1094428112452151.

[19] Glomsrud JA, Xie J. A Structured STPA Safety and Security Co-analysis Framework for Autonomous Ships. In: 29th European Safety and Reliability Conference; 2019.

[20] Hovden J, Albrechtsen E, Herrera IA. Is there a need for new theories, models and approaches to occupational accident prevention? Safety Science 2010;48(8):950–6. https://doi.org/10.1016/j.ssci.2009.06.002.

[21] Humayed A, Lin J, Li F, Luo B. Cyber-Physical Systems Security - A Survey. IEEE Internet of Things Journal 2017;4(6):1802–31. https://doi.org/10.1109/JIOT.2017.2703172.

[22] Johnson, B., Caban, D., Krotofil, M., Scali, D., Brubaker, N., & Glyer, C. (2017). Attackers Deploy New ICS Attack Framework "TRITON" and Cause Operational Disruption to Critical Infrastructure. Retrieved April 12, 2018, from https://www.fireeye.com/blog/threat-research/2017/12/attackers-deploy-new-ics-attack-framework-triton.html.

[23] Kavallieratos G, Katsikas S, Gkioulos V. Cybersecurity and Safety Co-Engineering of Cyberphysical Systems — A Comprehensive Survey. Future Internet 2020;12(65). https://doi.org/10.3390/fi12040065.

[24] Kavallieratos G, Katsikas S, Gkioulos V. SafeSec Tropos : Joint security and safety requirements elicitation. Computer Standards & Interfaces 2020;70(January):103429. https://doi.org/10.1016/j.csi.2020.103429.

[25] Kriaa S, Pietre-Cambacedes L, Bouissou M, Halgand Y. A survey of approaches combining safety and security for industrial control systems. Reliability Engineering & System Safety 2015;139:156–78. https://doi.org/10.1016/j.ress.2015.02.008.

[26] Kriaa S, Raspotnig C, Bouissou M, Pièrre-Cambacedes L, Karpati P, Halgand Y, Katta V. Comparing Two Approaches to Safety and Security Modelling : BDMP Technique and CHASSIS Method System architecture used for the case study. In: Proceedings of the Enlarged Halden Programme Group Meeting. Storefjell, Norway: OECD Halden Reactor Project; 2013.

[27] Langner R. Stuxnet: Dissecting a cyberwarfare weapon. IEEE Security and Privacy 2011;9(3):49–51. https://doi.org/10.1109/MSP.2011.67.

[28] Leveson NG. Safeware: System Safety and Computers. Addison-Wesley; 1995.

[29] Leveson NG. A new accident model for engineering safety systems. Safety Science 2004;42(4):237–70.

[30] Leveson NG. Engineering a safer world: systems thinking applied to safety. The MIT Press; 2011.

[31] Leveson, N. G. (2013). *An STPA Primer*.

[32] Leveson, N. G., & Thomas, J. P. (2018). *STPA Handbook*. https://doi.org/10.2143/JECS.64.3.2961411.

[33] Lyu X, Ding Y, Yang SH. Safety and security risk assessment in cyber-physical systems. IET Cyber-Physical Systems: Theory and Applications 2019;4(3):221–32. https://doi.org/10.1049/iet-cps.2018.5068.

[34] Paul S, Brunel J, Rioux L, Vallée F, de Oliveira J, Gailliard G, Chemouil D. Recommendations for security and safety co-engineering (Release No. 3). MeRgE ITEA2 Project; 2016.

[35] Pietre-Cambacedes L, Chaudet C. The SEMA referential framework: Avoiding ambiguities in the terms "security" and "safety. International Journal of Critical Infrastructure Protection 2010;3(2):55–66. https://doi.org/10.1016/j.ijcip.2010.06.003.

[36] Rasmussen B, Grønberg CD. Accidents and risk control. Journal of Loss Prevention in the Process Industries 1997;10(5–6):325–32. https://doi.org/10.1016/S0950-4230(97)00022-3.

[37] Raspotnig C, Karpati P, Katta V. A combined process for elicitation and analysis of safety and security requirements. Lecture Notes in Business Information Processing, Vol. 113. Springer Verlag; 2012. p. 347–61. https://doi.org/10.1007/978-3-642-31072-0_24. LNBIP.

[38] Rausand M. Risk Assessment: Theory, Methods, and Applications. John Wiley & Sons; 2011.

[39] Schmittner C, Ma Z, Puschner P. Limitation and Improvement of STPA-Sec for Safety and Security Co-analysis. A. Skavhaug, J. Guiochet, E. Schoitsch, & F. Bitsch (Eds.). Computer Safety, Reliability, and Security: SAFECOMP 2016 Workshops ASSURE, DECSoS, SASSUR, and TIPS. Trondheim; 2016. p. 195–209. https://doi.org/10.1007/978-3-319-45480-1. Retrieved from.

[40] Schmittner C, Ma Z, Schoitsch E, Gruber T. A case study of FMVEA and CHASSIS as safety and security co-analysis method for automotive cyber-physical systems. In: CPSS 2015 - Proceedings of the 1st ACM Workshop on Cyber-Physical System Security, Part of ASIACCS 2015; 2015. p. 69–80. https://doi.org/10.1145/2732198.2732204.

[41] Solberg CL. An STPA Analysis of the ReVolt. Norwegian University of Science and Technology; 2018.

[42] Stake RE. The Art of Case Study Research. SAGE; 1995.

[43] Sulaman SM, Beer A, Felderer M, Host M. Comparison of the FMEA and STPA safety analysis methods-a case study. In: *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft Fur Informatik (GI), P-292*; 2019. p. 175–6. https://doi.org/10.18420/se2019-55.

[44] Taylor Associates ApS. The QRAQ Project Quality of Risk Assessment for Process Plant - Volume 2, Quality and Completeness of Hazard Identification. Allerød; 2011.

[45] Taylor JR. Automated HAZOP revisited. Process Safety and Environmental Protection 2017;111:635–51. https://doi.org/10.1016/j.psep.2017.07.023.

[46] Wei LC, Madnick SE. A System Theoretic Approach to Cybersecurity Risk Analysis and Mitigation for Autonomous Passenger Vehicles. SSRN Electronic Journal 2019. https://doi.org/10.2139/ssrn.3370555. February.

[47] Weiss J. Protecting Industrial Control Systems from Electronic Threats. Momentum Press; 2010.

[48] Yampolskiy M, Horvath P, Koutsoukos XD, Xue Y, Sztipanovits J. Taxonomy for description of cross-domain attacks on CPS. In: Proceedings of the 2nd ACM International Conference on High Confidence Networked Systems - HiCoNS '13. 135; 2013. https://doi.org/10.1145/2461446.2461465.

[49] Young W, Leveson NG. Systems Thinking for Safety and Security. In: Proceedings of the 2013 Annual Computer Security Applications Conference (ACSAC); 2013. p. 1–8. https://doi.org/10.1145/2523649.2530277.

[50] Zio E. The Future of Risk Assessment. Reliability Engineering and System Safety 2018;177:176–90. https://doi.org/10.1016/j.ress.2018.04.020.