Doctoral theses at NTNU, 2021:198

Georgios Kavallieratos

Security of the Cyber Enabled Ship

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

NTNU
Norwegian University of Science and Technology

Georgios Kavallieratos

Security of the Cyber Enabled Ship

Thesis for the Degree of Philosophiae Doctor

Gjøvik, June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

NTNU

Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

ISBN 978-82-326-6790-1 (printed ver.)
ISBN 978-82-326-6979-0 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:198

Printed by NTNU Grafisk senter

Declaration of Authorship

I, Georgios Kavallieratos, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:

(Georgios Kavallieratos)

Date: March 2021

To Evaggelia,

Abstract

The maritime industry is actively engaged with developing remotely controlled and autonomous ships to sail in the near future. Remotely controlled and autonomous vessels have the potential to transform the maritime transport sector and to constitute the instantiation of the Industry 4.0 process in the maritime industry, termed "Shipping 4.0". Both remotely controlled and autonomous vessels are variants of the Cyber-Enabled Ship (C-ES), and comprise a number of interconnected Cyber Physical Systems (CPSs) that perform functions critical to the safe operation of the vessel. This proliferation of the use of integrated Information Technology and Operational Technology systems that aims to maximize the reliability and efficiency of a number of the vessel's operations, including vessel navigation, introduces previously unknown security risks that, in view of the significance of the sector to transportation and commerce, are important to address.

The overall objective of this research is to determine the security architecture of the C-ES seen as a system of CPSs, i.e. to provide a cohesive security design, which addresses the requirements - and in particular the risks of the C-ES, and specifies what security controls are to be applied where. Accordingly, the main research questions that the work described in this thesis addressed are as follows:

- What is a reference system architecture for the C-ES?
- What are the cyber security and safety risks and requirements of the C-ES?
- What is an appropriate security architecture for the C-ES?

In the course of addressing these research questions, we researched several aspects of the process of analyzing the security of CPSs and we proposed methods and approaches for carrying out such analysis. We thus effectively proposed a domain-agnostic approach for studying the security of complex interconnected CPSs, and we demonstrated its applicability to the case of the C-ES.

Specifically, we proposed methods for analyzing threats, attacks, attack paths, and risks of interconnected CPSs; for systematically selecting baseline security controls for individual CPSs; for eliciting security and safety requirements; and for selecting optimal sets of security controls for complex interconnected CPSs. We also proposed a reference architecture that can represent the C-ES in the maritime domain ecosystem, and a reference architecture for a cyber-physical range.

These results have been published in five journal articles and three articles in conference proceedings; these constitute the second part of the thesis.

Acknowledgements

This doctoral thesis is the outcome of my journey into academia that would never have been possible without the support of my supervisor Prof. Sokratis Katsikas, and of many others.

First and foremost, I am deeply grateful to my supervisor Prof. Sokratis Katsikas for his invaluable advice, continuous support and optimism, and patience during my PhD journey. He has always been available and willing to help, encourage, or even just relax my anxiety. I cannot thank him enough and I will always be greatly indebted.

I would like to thank my co-supervisors, Prof. Slobodan Petrovic, Prof. Edmund Førland Brekke, and Prof. Hao Wang for their support and suggestions.

Furthermore, I would also like to thank Dr Vasiliki Diamantopoulou, Dr Georgios Spathoulas, Dr Vasileios Gkioulos, and Dr Marios Anagnostopoulos who have been my closest co-workers and friends during my studies.

I hereby express my sincere thanks to the members of the evaluation committee, for agreeing to review and evaluate this PhD thesis.

I would like to thank all members of the CISaR group and the Department of Information Security and Communication Technology for the excellent working environment.

Last but not least, I would like to express my sincere gratitude to my family for bringing me up and more or less making me the person I am today. This journey became much happier with their support and encouragement.

Georgios Kavallieratos

Gjøvik 2021.

Contents

List of Figures vii

List of Tables viii

Abbreviations ix

Part I: Overview 1

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Background . . . 3
  - 1.2.1 The Cyber-Enabled Ship (C-ES) . . . 3
  - 1.2.2 Reference Architecture . . . 4
  - 1.2.3 Security Architecture . . . 5
- 1.3 Aim and Scope . . . 5
- 1.4 Summary of the results . . . 6
  - 1.4.1 List of publications . . . 6
  - 1.4.2 Additional Publications . . . 6
  - 1.4.3 Key results . . . 7
- 1.5 Thesis Structure . . . 8

2 The research problem 9
- 2.1 Related Work . . . 9
  - 2.1.1 System Architectures of the Cyber-Enabled Ship . . . 9
  - 2.1.2 Assessing and treating the security risks of the C-ES . . . 10
  - 2.1.3 Threat and attack modeling . . . 12
  - 2.1.4 Safety and Security Requirements engineering . . . 13
- 2.2 Research Questions . . . 14
- 2.3 Research methodology . . . 15

3 Overview of the research papers 17
- 3.1 Paper I: Modelling Shipping 4.0: A Reference Architecture for the Cyber-Enabled Ship . . . 18
- 3.2 Paper II: Managing Cyber Security Risks of the Cyber-Enabled Ship . . . 18
- 3.3 Paper III: Shipping 4.0: Security requirements for the Cyber-Enabled Ship . . . 19
- 3.4 Paper IV: Cybersecurity and safety co-engineering of cyberphysical systems: A comprehensive survey . . . 20
- 3.5 Paper V: SafeSec Tropos: Joint security and safety requirements elicitation . . . 20
- 3.6 Paper VI: Attack Path Analysis for Cyber Physical Systems . . . 21
- 3.7 Paper VII: Towards a cyber-physical range . . . 21
- 3.8 Paper VIII: Cyber risk propagation and optimal selection of cybersecurity controls for complex cyberphysical systems . . . 22

4 Conclusions, Limitations and Future work 23

4.1 Contributions 23
4.2 Limitations 24
4.3 Future work 25

Part II: Research Articles **39**

5 Article I: Modelling Shipping 4.0: A Reference Architecture for the Cyber-Enabled Ship [1] **40**

6 Article II: Managing Cyber Security Risks of the Cyber-Enabled Ship [2] **56**

7 Article III: Shipping 4.0: Security requirements for the Cyber-Enabled Ship [3] **76**

8 Article IV: Cybersecurity and Safety Co-Engineering of Cyberphysical Systems - A Comprehensive Survey [4] **86**

9 Article V: SafeSec Tropos: Joint security and safety requirements elicitation [5] **105**

10 Article VI: Attack Path Analysis for Cyber Physical Systems [6] **122**

11 Article VII: Towards a cyber-physical range [7] **138**

12 Article VIII: Cyber Risk Propagation and Optimal Selection of Cybersecurity Controls for Complex Cyberphysical Systems [8] **149**

List of Figures

1 The Design Science Research Methodology as a process 15

List of Tables

1 Abbreviations Table xi
2 Key results in published papers 8
3 Artifacts and key results 16
4 Articles and Research Questions 18

Abbreviations

Abbreviation	Description	Page
AI	Artificial Intelligence	1
C-ES	Cyber-Enabled Ship	1
ICT	Information and Communications Technology	1
MUNIN	Maritime Unmanned Navigation through Intelligence in Networks	2
AAWA	Advanced Autonomous Waterborne Applications	2
MAS	Mayflower Autonomous Ship	2
CPSs	Cyber-Physical Systems	2
CCS	China Classification Society	3
AIS	Automatic Identification System	4
ECDIS	Electronic Chart Display and Information System	4
GMDSS	Global Maritime Distress and Safety System	4
MAF	Maritime Architecture Framework	5
IIRA	Industrial Internet Reference Architecture	5
e-MAF	extended Maritime Architecture Framework	7
SGAM	Smart Grid Architectural Model	9
IMO	International Maritime Organization	9
MAXCMAS	Marine Autonomous Systems	9
SCADA	Supervisory control and data acquisition	10
ISM code	International Safety Management Code	10
MaCRA	Maritime Cyber-Risk Assessment	11
SRAM	A security risk analysis model	11
PLADD	Probabilistic Learning Attacker, the Dynamic Defender model	12
HAM	Hybrid Attack model	12
STRIDE	Spoofing, Tampering, Repudiation, Information disclosure, Denial of Service, Elevation of privileges	13
DREAD	Damage, Reproducibility, Exploitability, Affected Users, Discoverability	13
STPA	Systems Theoretic Process Approach	13
DSRM	Design Science Research Methodology	15
ICS	Industrial Control Systems	19
MASS	Maritime Autonomous Surface Ship	42
SCC	Shore Control Center	43
MarNIS	Maritime Navigation Information Services	44
BAS	Bridge Automation System	50
EAS	Engine Automation System	50
ERC	Exponential Ranking Centrality	52

BC	Betweenness Centrality	52
TD	Total Degree	52
IBS	Integrated Bridge System	60
INS	Integrated Navigational System	60
AEMC	Autonomous Engine Monitoring and Control System	61
EES	Engine Efficiency System	61
MIS	Maintenance Interaction System	61
NavS	Navigation Systems	61
ASC	Autonomous Ship Controller	61
HMI	Human-Machine Interface	61
RMSS	Remote Maneuvering Support System	61
EmH	Emergency Handling system	61
C.A.	Collision Avoidance	61
CCTV	Closed-circuit television	61
ASM	Advanced Sensor Module	61
AP	Auto Pilot	61
VDR	Voyage Data Recorder	61
EDL	Engine Data Logger	61
L	Low	61
M	Medium	61
H	High	61
VL	Very Likely	61
M	Moderate	61
R	Rare	61
GPS	Global Positioning System	82
HID	Human interface device	83
EFT	Extended Fault Tree	90
E-CFT	Extended Component Fault Tree	90
ASIL	Automotive Safety Integrity Level	91
SEL	Security Level	91
SSM	Six Step Model	91
HARA	Hazard Analysis and Risk Assessment	91
FACT	Failure-Attack-CounTermeasure	91
CRAF	Cyber Risk Assessment Framework	91
UFoI-E	Uncontrolled Flows of information and Energy	91
AVES	Automated Vehicles Safety and Security Analysis Framework	91
SCSD	Safety and Cybersecurity Deployment	91
DT	Defense Tree	92
SARA	Security Automotive Risk Analysis	92
AL	Autonomy Levels	106

ANS	Autonomous Navigation System	113
ETA	Estimated / Expected Time of Arrival	115
UCA	Unsafe Control Action	116
MADM	Multiple Attribute Decision Making	127
TIC	Tacit Input Centrality	127
TOC	Tacit Output Centrality	127
CC	Closeness Centrality	127
PMU	Phasor Measurement Units	140
PDC	Phasor Data Concentrator	140
EMS	Energy Management System	140
IED	Intelligent Electronic Devices	141
RTDS	Real Time Digital Simulator	141
VCSE	Virtual Control System Environment	141
VM	Virtual Machines	141
RICS	Resilient Information and Control Systems	142
AFTL	Analog Fluid Tank Lab	143
CAN	Control Area Network	143
ECU	Electronic Control Units	143
VCU	Vehicle Control Unit	143
SIEM	Security Information and Event Management system	146
SoS	System of Systems	150
GA	Genetic algorithms	153
TC	Total Cost	159

Table 1: Abbreviations Table

Part I: Overview

1 Introduction

This Chapter provides an overview of the PhD thesis. It aims to provide the motivation for the research project; to present the background necessary for the subsequent discussion; to describe the aim and scope of this thesis; to give an overview of the key results of the research work; and to outline the structure of the thesis. Details regarding the particular methods that were followed, along with the accordant results, are presented in detail in the research articles that constitute Part II of the thesis.

1.1 Motivation

There is intense activity of the maritime industry towards making remotely controlled and autonomous ships sail in the near future; this activity constitutes the instantiation of the Industry 4.0 process in the maritime industry, known as "Shipping 4.0". Industry 4.0 was initially coined to describe the coupling of Operational Technologies (OT) with Information Technologies (IT) and the application of contemporary Information and Communication Technologies (ICTs) to facilitate the monitoring and control of critical sectors [9, 10, 11]. Particularly, Industry 4.0 initially was meant to describe the trend towards the automation and data processing and storage technologies in manufacturing [12]. However, the term has been extended to encompass domains that are not usually considered as industrial, such as cities 4.0 [13] and banking 4.0 [14], where increased automation and connectivity are witnessed by leveraging technologies such as the Internet of Things, Artificial Intelligence (AI), and Big Data analytics. Accordingly, the term Shipping 4.0 was coined to describe the new developments in digitalization of shipping, to reflect developments similar to those in Industry 4.0 [11]. The new technological paradigm of Shipping 4.0 comes to alignment with Industry 4.0 with the development of remotely controlled and autonomous vessels [15] – both variants of the Cyber-Enabled Ship (C-ES). Such vessels are equipped with autonomous and advanced decision support systems that facilitate the monitoring and control of the integrated OT/IT systems on board.

The concept of an unmanned ship is not new; visions of such ships were reported as early as the 1970's and have continued to appear regularly since then [16]. In 1973, Schönknecht et al [17] stated their vision of the unmanned ship. In 1980, the Japanese Intelligent Ship project aimed at "bringing about 'intelligent ships' that can function without help from the crew" and proposed the Master-Slave ship model, according to which robot ships would form convoys. In 1994, Kai Levander proposed the "Ship without crew" for short-sea shipping. In 1996, Bertram and Kaeding suggested that a combination of AI and teleoperation was feasible for ships. Back then, however, the concept was still not attractive to shipping com-

panies due to high maintenance costs. The concept re-emerged in a 2007 paper on the future development of the maritime industry by Waterborne TP, a cluster of European maritime stakeholders. Although this paper suggested that more advanced automation and improved sensors might be desirable, it stopped short of advocating full automation. Five years later, inspired by this idea, European research groups launched the collaborative MUNIN (Maritime Unmanned Navigation through Intelligence in Networks) project, co-funded under the EU's FP7 research programme. MUNIN concluded that an unmanned and autonomous merchant vessel was possible, but only for deep-sea voyage and not in congested or restricted waters, where a crew should operate the ship [18]. In 2013, Oscar Levander of Rolls-Royce, proposed the building of "unmanned container ships", by leveraging a combination of AI and tele-operation. In 2014, DNV GL proposed the concept of ReVolt, a 60-meter long, battery powered, unmanned container feeder vessel to sail the territorial waters of Norway [19]. In 2016, Rolls-Royce initiated a joint industry project in Finland called Advanced Autonomous Waterborne Applications (AAWA) to create the technology for a remotely controlled or fully autonomous ship to operate in coastal waters [20].

Today, many companies, mostly from Scandinavian countries, Korea, and Japan, are working on full-size autonomous ships with the goal of obtaining cargo vessel or even passenger vessel capabilities, and relevant major research projects are underway [21]. Yara Birkeland, the world's first fully autonomous and electric container vessel is expected to be ready to sail in 2022 [22]. In 2020, Ocean research non-profit ProMare and IBM announced the completion and launch of the Mayflower Autonomous Ship (MAS) – an AI and solar powered marine research vessel which will traverse oceans gathering vital environmental data [23]. The concept and prototype of an autonomous all-electric passenger ferry for urban water transport have been developed by the Norwegian University of Science and Technology within the ongoing Autoferry project [24].

The new technological era in the maritime domain aims to unify embedded systems with communication technologies that interact with the ship's physical environment to address economical, environmental, and safety issues; such systems are referred to as Cyber-Physical Systems (CPSs). The increased automation and autonomy minimize the costs of the crew and port handling, and at the same time minimize the safety risks that are caused by human error, such as collisions and environmental pollution.

On the other hand, the increasing adoption of interconnected CPSs into the vessel's networks increases the attack surface and the sophistication of potential cyber attacks. Five types of attackers, with varying accessibility, capability, and motivation profiles have been identified in the maritime sector: activists, competitors, criminals, terrorists, and elitists [25]. Indeed, various cyber attacks have been launched in the past few years, targeting either shore-based infrastructures (e.g. ports) [26] or onboard systems (e.g. navigational systems) [27, 28]. Cybersecurity incidents with particular focus on cyber attacks on bridge and navigational systems have been studied in [29]. The main types of attacks against vessels are jamming [30], spoofing, and hijacking, whilst the main type of attack against port infrastructures is data loss [31, 32]. The existence of legacy systems, where security by design principles have not been considered, and the lack of security awareness on board are some of the major issues that allow cyber-attacks to succeed [33].

The advantages of the digitalization process notwithstanding, the interconnectivity of the vessels' CPSs increases the attack surface and poses significant security risks that need to

be managed. Considering that the worldwide trade highly depends on the shipping industry [34], the identification and mitigation of the risks that the vessels face should be done early in the design phase so that the vessels are effectively fended against different types of attackers and sophisticated cyber attacks.

It is not surprising, therefore, that the importance of security for cyber-enabled vessels has been highlighted in technical reports and guidelines [35, 36]. Additionally, the importance of a cyber security framework to ensure the necessary levels of security, especially in vessels with increased autonomy levels, has been emphasized in the research roadmap for smart and autonomous maritime transport systems of [37]. Hence, the study of the security of cyber-enabled vessels is important, and timely. To this end, security risks need to be identified and assessed, threats and attacks need to be modeled and analyzed, and security and safety requirements need to be elicited, towards defining a security architecture for the cyber-enabled vessels.

1.2 Background

In this section we discuss key concepts that underpin the work described in the thesis. The concepts of the C-ES, of the reference architecture, and of the security architecture are discussed.

1.2.1 The Cyber-Enabled Ship (C-ES)

A Cyber-Enabled system is a computerized, automated, or autonomous system that is characterized by logic, data processing hardware, processing hardware, behavior governing software, and external communications capabilities [38]. The term "C-ES" was coined by Lloyds's Register in one of the first attempts to capture the integration of interconnected and automated systems in the maritime context, and particularly in the vessels' infrastructure. The concept of the C-ES encompasses different types of vessels, with varying levels of automation: vessels with automated processes and decision support systems; remotely controlled vessels with or without people onboard; and fully autonomous vessels with advanced support decision systems [39]. Lloyd's Register provided guidance for C-ESs by analyzing the integration of cyber systems into a ship's infrastructure, along with their technical and managerial considerations [40, 41]. In addition, DNV-GL provided guidance towards establishing novel systems onboard autonomous and remotely controlled vessels without compromising safety and secure navigation [35]. The regulatory and the technological landscapes of autonomous vessels are also analyzed by DNV-GL in [42], and the China Classification Society (CCS) proposed general guidelines for autonomous cargo ships [43].

Although the increasing adoption of new technologies in shipping enables the realization of autonomous and remotely controlled vessels, performance and security issues may arise from the adoption of Industry 4.0 technologies, and in particular by the high automation and interconnectivity. The C-ES is a cyber-physical ecosystem that includes critical cyber-physical systems. Such systems are characterized by functional and operational requirements that aim to ensure the safe and secure operation of the vessel. Operational requirements of

critical control systems have been identified in [44]. These are "real-time performance", "dependability", "sustainability", "survivability", and "safety-critical". Real-time performance pertains to the system's ability to operate in real (or near-real) time, so as to be able to meet certain deadlines associated to the controlled process. Dependability refers to the ability of the CPS to execute a service on time, by avoiding frequent and severe internal faults [45]. Sustainability is defined as the ability of a system to meet the current need without compromising the ability of future developments to meet their own needs [10]. Four attributes of sustainability have been defined in [44]; scalability, extensibility, interoperability, and maintainability. Survivability is defined as the ability of the system to accomplish its mission and deal with malicious, deliberate or accidental faults in time [46]. The safety-critical requirement is a variant of safety for critical environments and describes systems that may lead to catastrophic consequences due to faults, and which could cause human losses or injuries, or physical damage to the infrastructure [47]. The aforementioned requirements, particularly the requirements for sustainability and scalability of critical systems, increase the need for thorough security and safety analysis of the CPSs of the C-ES and the identification of the accordant security and safety requirements.

The C-ES ecosystem consists of the vessel itself, a Shore Control Center (SCC), and other vessels in the vicinity. These are interconnected by means of telecommunication systems. The vessel itself comprises the bridge, engine, and IT systems, whilst the SCC includes systems that enable the monitoring and control of the vessel. The higher the autonomy level of the vessel, the more information is exchanged among systems and more advanced systems and services are introduced. Consequently, the operations of remotely controlled and autonomous vessels are the same with those of conventional ships, but technical services, control systems, information exchanged, and functions of the vessel are different; a remotely controlled or autonomous vessel makes use of a larger number of services, and has more CPSs on board. Conventional vessels are equipped with navigational systems such as the Automatic Identification System (AIS), the Electronic Chart Display and Information System (ECDIS), and the Global Maritime Distress and Safety System (GMDSS). In addition to those, key CPSs of the C-ES's bridge systems are the autonomous navigational system, autonomous collision avoidance systems, and autonomous or remotely controlled maneuvering systems. Further, autonomous and remotely controlled vessels highly rely on information of sensors and actuators, since autonomous decision making systems demand high information accuracy. The autopilot, weather sensors, and environmental analysis sensors are important for conventional ships; additionally, engine actuators, navigation and docking sensors are important for the C-ES. Therefore, inasmuch a conventional ship's infrastructure needs to accommodate simple vessel functions, such as autopilot and weather observations, C-ESs will be equipped with advanced sensor systems, able to facilitate functions such as docking, mooring, and engine maintenance.

1.2.2 Reference Architecture

A reference architecture outlines restrictions for an instantiation - concrete architecture - by describing the structure of a system. The system element types, structures, along with their interaction types among each other and with their environment constitute a reference

architectural model. A reference architecture facilitates the representation of systems under development, whose individual details have not been established yet; this representation is achieved by means of abstraction layers that characterize reference architectural models. Based on these models, further architectural instances can be developed for ecosystems with the same functional requirements. In addition, a reference architecture provides the means to understand and select effective methods; to develop durable architectures and specific designs; and to create simulation models [48]. A reference architecture is derived by the combination of the conceptual model of a system with its characteristics that develop reference models to describe architectural instantiations [49].

Several examples of reference architectures in different domains exist: the Maritime Architecture Framework [50]; the RAMI 4.0 Reference Architectural Model for Industrie 4.0 [51]; the Industrial Internet Reference Architecture (IIRA) [52]; several reference architecture models for digital manufacturing [53]; the International Organization for Standardization reference architecture for the Internet of Things [49]; a number of other reference architectures for the Internet of Things [54]; a security reference architecture for IoT-based smart homes [55]; reference architecture models for smart cities [56, 57]; and the Smart Grid Reference Architecture [58].

1.2.3 Security Architecture

The definition of a security architecture is not consistent in the literature. Some view it as a framework, others as a process, yet others as a detailed technical design. Herewith, we adopt the definition of the Swiss Information Security Society, according to which the security architecture is "a cohesive security design, which addresses the requirements – and in particular the risks of a particular environment/scenario, and specifies what security controls are to be applied where. The design process should be reproducible".

The word 'architecture' is used at all levels of detail within a system. Following [59], here we are concerned with the high-level design of a system from a security perspective, in particular how the primary security controls are motivated and positioned within the system. The same view has been adopted in [60].

1.3 Aim and Scope

This research aims to define the security architecture of the remotely controlled and of the autonomous vessel, both variants of the Cyber-Enabled Ship.

Defining the security architecture of the C-ES is a multifaceted objective, that breaks down into a number of secondary, supportive objectives. These are the analysis of security risks, threats, and attacks; and, based on these, the identification of the security requirements; and the subsequent selection of the appropriate security controls. Due to the strong coupling between the cyber and the physical world that cyber physical systems bring, safety needs also to be considered. In this research, the analysis is at the system level; specific implementation details, such as e.g. communication protocols, or specific vessel types are not considered.

1.4 Summary of the results

This thesis comprises the results included in eight published articles that are listed in section 1.4.1. Within the course of the research project, five additional articles were published, which are not included in the thesis; these are listed in section 1.4.2. An overview of the key results of this research is given in section 1.4.3.

1.4.1 List of publications

1. *Paper I:* Kavallieratos, Georgios; Katsikas, Sokratis; Gkioulos, Vasileios. (2020). "Modelling Shipping 4.0: A Reference Architecture for the Cyber-Enabled Ship." Asian Conference on Intelligent Information and Database Systems. Springer, Cham, 2020, Part II, LNCS 12034 proceedings.
2. *Paper II:* Kavallieratos, Georgios, and Sokratis Katsikas. "Managing Cyber Security Risks of the Cyber-Enabled Ship." Journal of Marine Science and Engineering 8.10 (2020): 768.
3. *Paper III:* Kavallieratos Georgios, Diamantopoulou Vasiliki, Katsikas Sokratis. (2020) Shipping 4.0: Security requirements for the Cyber-Enabled Ship. IEEE Transactions on Industrial Informatics.
4. *Paper IV:* Kavallieratos, Georgios; Katsikas, Sokratis; Gkioulos, Vasileios. (2020) Cybersecurity and safety co-engineering of cyberphysical systems: A comprehensive survey. Future Internet, 2020, 12(4), 65.
5. *Paper V:* Kavallieratos, Georgios; Katsikas, Sokratis; Gkioulos, Vasileios. (2020) SafeSec Tropos: Joint security and safety requirements elicitation. Computer Standards & Interfaces. vol. 70.
6. *Paper VI:* Kavallieratos, Georgios and Katsikas, Sokratis. (2020). "Attack Path Analysis for Cyber Physical Systems." Computer Security ESORICS 2020 International Workshops, CyberICPS 2020 and SECPRE 2020, Surrey, UK, September 14–18, 2020.
7. *Paper VII:* Kavallieratos, Georgios; Katsikas, Sokratis; Gkioulos, Vasileios. (2019) "Towards a Cyber-Physical Range". In Proceedings of the 5th on Cyber-Physical System Security Workshop (pp. 25-34).
8. *Paper VIII:* Kavallieratos, Georgios; Spathoulas Georgios; Katsikas, Sokratis. (2021) Cyber risk propagation and optimal selection of security controls for complex cyber-physical systems. Sensors, 2021, 21(5):1691.

1.4.2 Additional Publications

These publications contribute to security research, particularly in cyber-physical systems, Internet of Things, information system networks security, and maritime security.

1. Katsikas, Sokratis; Kavallieratos, Georgios. (2021) Cybersecurity of the unmanned ship, Chapter in Cybersecurity Issues in Emerging Technologies. CRC Press, 2021.
2. Kavallieratos, Georgios; Gkioulos, Vasileios; Katsikas, Sokratis. (2019) "Threat Analysis in Dynamic Environments: The Case of the Smart Home". Proceedings of the 15th Annual International Conference on Distributed Computing in Sensor Systems - DCOSS 2019.
3. Kavallieratos, Georgios; Chowdhury, Nabin; Katsikas, Sokratis; Gkioulos, Vasileios; Wolthusen, Stephen. (2019) Threat Analysis for Smart Homes. Future Internet. vol. 11 (10).
4. Belalis, Ilias; Kavallieratos, Georgios; Gkioulos, Vasileios; Spathoulas, Georgios. (2020) "Enabling Defensive Deception by Leveraging Software Defined Networks". The Twelfth International Conference on Evolving Internet INTERNET 2020
5. Ahmed Amro; Kavallieratos, Georgios; Louzis, Konstantinos; Thieme, Christoph. (2020) "Impact of cyber risk on the safety of the MilliAmpere2 Autonomous Passenger Ship". Proceedings of The third International Conference on Maritime Autonomous Surface Ship (ICMASS) 2020.

1.4.3 Key results

The key results of this research work are as follows:

- The extended Maritime Architecture Framework (e-MAF) that describes aspects of the C-ES ecosystem, and the C-ES CPS architecture.
- Modified STRIDE and DREAD methods for analyzing threats and security risks in CPSs.
- A systematic approach to select baseline security controls to treat the security risk of CPSs.
- A security requirements elicitation process for CPSs.
- A multi attribute taxonomy of security and safety co-engineering methods.
- A method for the joint elicitation of safety and security requirements for CPSs.
- A method for discovering and analyzing attack paths within interconnected CPSs.
- A reference model for a cyber physical range.
- A method for assessing the aggregate security risk of large scale, complex CPSs, comprising interconnected components.
- A method for selecting an optimal set of security controls among those in an established knowledge base, that reduce the residual risk while at the same time minimizing the cost.
- Security controls for the navigational systems of two instances of the C-ES, namely the remotely controlled ship and the autonomous ship derived by employing the methods

Articles	Key results
Paper I	e-MAF; C-ES CPS architecture.
Paper II	Modified STRIDE and DREAD methods; systematic approach for selecting security baseline controls.
Paper III	A security requirements elicitation process for CPSs.
Paper IV	A multi attribute taxonomy of security and safety co-engineering methods.
Paper V	A method for the joint elicitation of safety and security requirements for CPSs.
Paper VI	A method for discovering and analyzing attack paths within interconnected CPSs.
Paper VII	A reference model for a cyber physical range.
Paper VIII	A method for assessing the aggregate security risk of large scale, complex CPSs, comprising interconnected components; a method for selecting an optimal set of security controls among those in an established knowledge base, that reduce the residual risk while at the same time minimizing the cost; Security controls for the navigational systems of two instances of the C-ES, namely the remotely controlled ship and the autonomous ship derived by employing the methods developed within this research.

Table 2: Key results in published papers

developed within this research.

The mapping between published articles and the above key results is shown in Table 2.

1.5 Thesis Structure

This thesis is organized into two parts. The first part consists of four chapters and gives an overview of the overall research project. Chapter 1 is this introduction. Chapter 2 describes the research problem. It includes an overview of the related work, that leads to the research questions that the thesis addresses. It also includes the research methodology and the methods that were employed to address particular research questions. Chapter 3 presents a summary of the published articles that make up the second part of the thesis. Chapter 4 is the conclusion. It includes the contributions of this research; limitations of the research; and insights into future work. The second part of the thesis includes the eight peer reviewed articles that built the basis for this research and constitute the main part of the thesis.

2 The research problem

2.1 Related Work

In this section we review existing relevant literature that constitutes the baseline for the work in the thesis, structured into broad themes. Detailed literature reviews on the specific topics that were examined in the research are included in the articles in Part II. The goal is to identify limitations and to recognize research gaps to be addressed in this thesis. We start with reviewing works that describe the systems, architectures, functions, and operations of the autonomous ships. Subsection 2.1.2 reviews the related work on assessing and treating security risks of the C-ES, whilst subsection 2.1.3 reviews threat and attack modeling techniques. Further, subsection 2.1.4 reviews safety and security requirements engineering approaches.

2.1.1 System Architectures of the Cyber-Enabled Ship

The maritime industry is making fast progress towards the development of remotely controlled and autonomous vessels. The establishment of a reference architecture facilitates the aforementioned progress by synthesizing and analyzing both legacy and new technologies that co-exist on the vessel. A Maritime Architecture Framework (MAF) was proposed in [50], to facilitate the development and adoption of new systems and technologies in the maritime domain. The development process of the MAF followed that of the Smart Grid Architectural Model (SGAM) [58]; accordingly, the MAF has been developed taking into consideration existing maritime architectures, including the Common Shore Based System Architecture [61] and the International Maritime Organization's (IMO) e-Navigation architecture [62]. However, the MAF cannot accommodate autonomous or remotely controlled ships.

Several works in the literature have proposed system architectures for autonomous and remotely controlled vessels; however, all of these focus on specific vessels (e.g. merchant, pallet shuttle barge), and systems (e.g. communication systems, collision avoidance systems). An autonomous navigation system and technologies for path planning and collision avoidance were proposed within the AAWA project [20]. An ICT architecture of unmanned merchant ships was proposed in [63], and the communication architecture was described in [64]. The MUNIN project published several deliverables that analyzed the architectures and the operations of the bridge [65], the Shore Control Center [66] and the engine rooms [67] of an unmanned merchant ship. A generic system architecture for the collision avoidance system of an autonomous ship was developed within the Marine Autonomous Systems (MAXCMAS) project [68]. The core systems of an autonomous pallet shuttle barge were proposed in [69] and the functional requirements of such systems were described in [70].

A hierarchical structure for the systems of autonomous cargo ships was proposed in [71]. The key technologies that contribute towards the development of autonomous surface ships, particularly focusing on the vessel, the control center, and the communication infrastructures

were discussed in [72]. A generic architecture for unmanned vessels based on Intelligent Information Technology was proposed in [73]. However, the analysis focused on the operations and future technologies; a system architecture was not proposed. An architecture focusing on both satellite and terrestrial communication systems was proposed in [74]. An architecture focusing on the situational awareness system for autonomous and remotely controlled vessels was proposed in [75].

2.1.2 Assessing and treating the security risks of the C-ES

Guidelines for managing cybersecurity risks in the maritime sector have been proposed by the IMO and by maritime classification societies. The IMO provided high level recommendations on maritime cyber risk management based on international standards and on the existing International Safety Management (ISM) code [76, 77]. Further, general guidance for cyber security management both for onboard and for shore side systems is provided in [36]. General security threats in the maritime sector have been discussed in [78], where also high level recommendations for a systematic security assessment in conventional maritime systems are provided. A code of practice for cybersecurity onboard has been developed in [79] to ensure the cyber security resilience onboard conventional vessels. Additionally, general security requirements and measures for Informational Technology onboard have been proposed in [80]. However, the aforementioned reports and guidelines focus on conventional ships and provide only general frameworks for cyber risk management. As the CPSs encountered in the C-ES are characterized by high interconnectivity and autonomy, traditional frameworks such as the aforementioned are rarely adequate [36].

Risk assessment is a sub-process of the risk management process. Risk assessment methodologies enable the identification, analysis, and evaluation of the security risks. Many security risk assessment methods applicable to general purpose IT systems exist [81], and a number of taxonomies and comparison frameworks have been proposed to classify and analyse them [82, 83, 84]. Even though several of these methods can be and have been applied to CPSs, they cannot accurately assess cyber risks related to CPSs [85].

A number of approaches for risk assessment for CPSs published before 2015 appeared in [85]. A more recent review of a few risk assessment methods for CPS, from the perspective of safety, security, and their integration, including a proposal for some classification criteria was made in [86]. Cyber risk assessment methods for CPSs more often than not are domain specific, as they need to take into account safety as an additional impact factor. Overviews of such domain-specific methods for the smart grid, the Internet of Things, Supervisory control and data acquisition (SCADA) systems, and the automotive domain were provided in [87], [88], [89], and [90] respectively.

Even though the cyber security of other modes of autonomous transport, such as vehicles and railways, has been extensively researched, the cyber security risks of the C-ES have only been examined and analyzed scarcely. General cyber attacks that pose risks for autonomous ships, along with the potential controls to mitigate such risks were discussed in [91]. A cyber risk assessment methodology to analyze cyber risks of autonomous ships and the potential cyber attacks from the attacker's perspective was proposed in [92], and a model-based risk assessment framework called MaCRA (Maritime Cyber-Risk Assessment) was proposed in

[25]. General cyber risks of autonomous ships assessed using MaCRA were discussed in [93]. A risk assessment of the navigational and propulsion systems of an inland vessel was provided in [94]. Additionally, general cyber security issues of autonomous ships were discussed in [38, 95]. Although this framework provides a comprehensive picture of the maritime risk and the factors that may influence it, details on the inherent risk of each component and on how the risk propagates between interconnected components are missing.

The C-ES variants are systems still under development. For such systems, whose operational and functional requirements have not yet been established, risk assessment is best performed by means of a combination of qualitative and quantitative methods, so as to obtain a holistic view. Additionally, such a hybrid approach facilitates the communication of the results to relevant stakeholders while allowing the representation of cyber risk in numeric form, thus facilitating the assessment of the effectiveness of controls at later stages of the risk treatment process. An approach to risk assessment of highly interconnected CPSs comprising heterogeneous components with the aforementioned characteristics is yet to be proposed and applied to the case of the C-ES.

Several works in the literature have studied how individual elements of the security risk (threats, vulnerabilities, impacts) propagate in a network of interconnected systems; both deterministic and stochastic approaches have been used to this end. A threat likelihood propagation model for information systems based on the Markov process was proposed in [96]. An approach for determining the propagation of the design faults of an information system by means of a probabilistic method was proposed in [97]. A security risk analysis model (SRAM), based on a Bayesian network, that allows the analysis of the propagation of vulnerabilities in information systems was proposed in [98]. Methods for evaluating the propagation of the impact of cyber attacks in CPSs were proposed in [99, 100, 6], among others. Epidemic models were initially used to study malware propagation in information systems [96]. The propagation of security incidents in a CPS was viewed as an epidemic outbreak in [101] and it was analyzed using percolation theory. The method was shown to be applicable for studying malware infection incidents, but it is questionable whether the epidemic outbreak model fits other types of incidents. Percolation theory was also used in [102] to analyze the propagation of node failures in a network of CPSs comprising cyber and physical nodes organized in two distinct layers, such as in the case of the power grid. A quantitative risk assessment model that provides asset-wise and overall risks for a given CPS and also considers risk propagation among dependent nodes was proposed in [103].

A method for assessing the aggregate risk of a set of interdependent critical infrastructures was proposed in [104, 105]. The method provides an aggregate cyber risk value at the infrastructure level, rather than a detailed cyber risk assessment at the system/component level. Thus, it is suitable for evaluating the criticality of infrastructure sectors, but not for designing security architectures or for selecting appropriate security controls. A similar approach for the Energy Internet [106] was followed to develop an information security risk algorithm based on dynamic risk propagation in [107]. A framework for modeling and evaluating the aggregate risk of user activity patterns in social networks was proposed in [108]. A two-level hierarchical model was used in [109] to represent the structure of essential services in the national cyberspace, and to evaluate the national level (aggregate) risk assessment by taking into account cyber threats and vulnerabilities identified at the lower level. Therefore, risk propagation among and aggregation at components in a CPS has not been adequately

researched.

Risk treatment is another sub-process of the risk management process that aims to select the appropriate security controls to minimize, retain, avoid, and/or share the assessed security risks. The security controls must satisfy the established security requirements and lead to the development of a security architecture. The methods reviewed above mainly focus on the analysis and assessment of security risks, and only partially address the risk treatment sub-process.

The systematic selection of the most appropriate security controls that will lead to the security architecture of a CPS has been only partially studied in the literature [110, 111]. The selection of the security controls is still largely performed empirically, particularly for CPSs.

2.1.3 Threat and attack modeling

Threat and attack modeling techniques enable the comprehensive study of cyber threats and attacks by analyzing the adversary's profile, the goals of the attack, the techniques used to launch it, and the sequence of events that lead to a successful cyber attack. Both methods are important instruments towards gaining insight into cyber attacks; and both are essential in the process of identifying appropriate security controls.

Many threat modeling methodologies for ICT systems have been proposed in the literature. Several of these methods have been surveyed and their key characteristics have been identified in [112, 113, 114]. Similarly, several cyber attack analysis techniques for ICT systems have been proposed in the literature [115, 116].

A survey of attack modeling methods in cyber physical domains was performed in [117]. The Markov Chain Model, the Probabilistic Learning Attacker, the Dynamic Defender (PLADD) model, and the Hybrid Attack model (HAM) were reviewed in [118]. This survey focused on hybrid models that provide a more comprehensive view of attacks and the accordant security defenses. A review of the Graph-based method, the Bayesian network-based method, the Markov model-based method, the cost optimization method, and uncertainty analysis was conducted in [119]. A survey of methods for assessing attack paths to critical infrastructures and services was performed in [120]. Although this survey considers Internet of Things, there is a very extended part that focuses on CPS-based environments. Attack trees and graphs have been extensively utilized to analyze interconnected systems and attack paths between such systems. Their main advantage over other types of attack models is that they can identify all possible attacks on a system. However, a major disadvantage is that these methods do not scale well.

Threat and attack modeling in autonomous and remotely controlled vessels is yet to be comprehensively analyzed. General security threats for the navigational, propulsion, and cargo-related systems have been discussed in [27]. This study focused on potential attack scenarios considering existing vulnerabilities. However, the analysis targeted conventional vessels with cyber capabilities, and did not follow a systematic process to identify attack scenarios. Early studies of cyber attacks against autonomous ships appeared in [121, 70, 95].

Considering the existing threat and attack modeling methods, STRIDE[1], DREAD[2] and attack graphs are selected as the most appropriate to analyze highly interconnected CPSs that comprise heterogeneous components [124, 125, 126]. In particular, the interrelated STRIDE and DREAD methods provide a comprehensive understanding of cyber attacks; DREAD facilitates the rating, comparison, and prioritization of the severity of STRIDE threats and provides a flexible scoring approach that can be extended to incorporate CPSs aspects. Moreover, STRIDE and DREAD can analyze systems whose detailed operational and functional requirements have not been yet established, in contrast to other approaches that need such requirements to produce valid results [113]. As such, they are appropriate for use in systems still at the development stage.

2.1.4 Safety and Security Requirements engineering

Several security requirements elicitation methods for ICT systems have been proposed in the literature and have been reviewed in relevant surveys [127, 128, 129]. Among these, Secure Tropos [130] extends the Tropos [131] method so as to enable the capturing of security aspects, and combines requirements engineering concepts, such as "actor," "goal," and "plan," together with security engineering concepts such as "threat," "security constraint," and "security mechanism". Further, various approaches for security analysis based on Secure Tropos have been proposed in the literature [132, 133, 134]. The Secure Tropos methodology has been recommended in several surveys [135, 136] as an appropriate method for analyzing systems under development.

Likewise, several safety requirements analysis methods have been proposed in the literature and have been reviewed in [137, 138]. Among them, the Systems Theoretic Process Approach (STPA) is a prominent systematic safety analysis method that focuses on the control actions of the targeted system [139]. The advantages of STPA as compared to other alternatives are the wider perspective it provides on the system hazards; its ability to capture the control structure; and its coverage of conflicting actions in CPSs. Various variants of the STPA have been proposed in the literature [140, 141].

A systematic literature review of methods for the joint analysis of safety and security was conducted in [142], and several safety and security co-analysis methods were reviewed in [86, 143]. A comprehensive survey of safety and security co-engineering methods was conducted in [144]. In this survey, existing approaches were classified according to whether they are graphical or non-graphical, and whether they follow a unified or integrated approach to combine safety and security. Further, a survey of the existing safety assurance methods able to analyze CPSs was carried out in [137]. Most of the existing methods are unified approaches that lead to incomplete results [145], particularly as they more often than not result in conflicting requirements. A framework able to detect conflicts between safety and security requirements early in the development phase was proposed in [146]. The conflict resolution between safety and security requirements based on the NIST SP 800-30 method and the STPA was proposed in [147].

[1]STRIDE stands for Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privileges [122].

[2]DREAD stands for Damage, Reproducibility, Exploitability, Affected Users, and Discoverability [123].

The security requirements of autonomous vessels have only been scarcely and non-systematically examined: The technical and non-technical communication requirements for an autonomous merchant ship were analyzed in [64]. The data requirements for wireless transmission of autonomous ships were identified in [148]. The functional requirements of six main systems of the autonomous ship were presented in [149]. The security requirements for conventional vessels were described in [150].

Likewise, the joint security and safety requirements for autonomous and remotely controlled vessels have been also studied scarcely in the literature. General security and safety aspects of maritime vessels were discussed in [151]. The security and safety issues of a semi-autonomous vessel were analyzed in [152]. Further, a method to combine security and safety risks of the collision avoidance function of an autonomous surface vessel was presented in [153]. Autonomous ships have been used to illustrate the workings of various co-analysis methods [154, 155, 156]. However, a systematic analysis of safety and security requirements of the C-ES ecosystem and its constituent CPSs is yet to be developed.

2.2 Research Questions

The overall objective of this research is to determine the security architecture of the C-ES, i.e. to provide a cohesive security design, which satisfies the requirements and manages the risks of the C-ES, and specifies what security controls are to be applied where.

In order to reach this objective, in view of the research gaps identified in the previous section, this research is driven by the following research questions and sub-questions:

- **Research question 1:** What is a reference system architecture for the C-ES?
  - ***Research question 1.1:*** What cyber-physical systems make up the C-ES?
  - ***Research question 1.2:*** What are the interconnections and interdependencies among the cyber-physical systems making up the C-ES?
- **Research question 2:** What are the security and safety risks and the accordant requirements of the C-ES?
  - ***Research question 2.1:*** How can we assess the combined security and safety risks of the C-ES?
  - ***Research question 2.2:*** How can we identify security and safety requirements for the C-ES's cyber-physical systems?
  - ***Research question 2.3:*** What are the threats that the C-ES faces, what cyber attacks can exploit these threats, and how can these be analyzed and modeled?
  - ***Research question 2.4:*** What is the architecture of a testbed for testing the security of the C-ES?
- **Research question 3:** What is an appropriate security architecture for the C-ES?

– ***Research question 3.1:*** How can the security and safety risks of the C-ES be treated?

2.3 Research methodology

Overall, this research was guided by the Design Science Research Methodology (DSRM) [157]. The DSRM is widely used to produce systems that are under development, such as the C-ES, by modifying existing situations to achieve better results [158]. The method enables the exploration, description, and explanation of a research problem along with the design and subsequent evaluation of the appropriate solutions (artifacts) for it [158]. In the particular case of information systems, these artifacts have been classified into eight categories: System Design (A description of an IT-related system); Method (Define the activities to create or interact with a system); Language/Notation (A -generally formalized- system to formulate statements that represents parts of reality); Algorithm (An executable description of behavior of a system); Guideline (Provide a generalized suggestion about system development); Requirements (Statements about a system); Pattern (Definition of reusable elements of design with its benefits and application context); and Metric (A mathematical model that is able to measure aspects of systems or methods) [159]. In this research we designed and developed artifacts in the System Design; Methods; Algorithms; Requirements; Metric; and Guidelines classes. The mapping between the key results of the thesis and artifacts developed within this research is depicted in Table 3.

The DSRM as a process is depicted in Figure 1; it consists of six activities: i) Identify the problem, ii) Define the objectives for the solution, iii) Design and development, iv) Demonstration, v) Evaluation, and vi) Communication [159]. DSRM is an iterative process, and can be initiated at any of its stages.

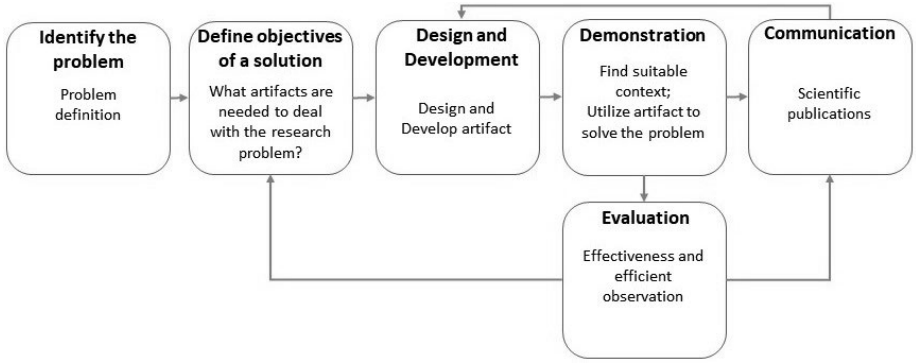


Figure 1: The Design Science Research Methodology as a process

Although the DSRM guided the overall research process, additional research methods, appropriate for security research [160], were employed when addressing particular research questions. The case study method [158] was employed when addressing RQ1, to illustrate the workings of the proposed e-MAF in three instances of the C-ES, namely the conventional,

Artifacts	**Results**
Develop System Design	eMAF; C-ES CPS architecture; cyber physical range reference architecture.
Methods	Security requirements elicitation process; a joint elicitation of safety and security requirements for CPSs; a method for discovering and analyzing attack paths in CPSs; a method for assessing the aggregate risk in complex CPSs; an approach to select the security baseline controls for individual CPSs; a method to select an optimal set of security controls.
Algorithms	An algorithm to analyze the aggregate risk in complex and interconnected CPSs; an algorithm to select the optimal set of security controls for complex and interconnected CPSs.
Requirements	General and system specific security requirements for C-ES CPS; safety and security requirements for the most vulnerable CPSs of the C-ES.
Guidelines	A set of baseline security controls for the most vulnerable CPSs of the C-ES; two optimal sets of security controls for the CPSs of the autonomous and remote controlled vessels.

Table 3: Artifacts and key results

the remotely controlled, and the autonomous vessel. The literature review method [161] was employed in addressing RQ2 and RQ3; both systematic and semi-systematic reviews were employed. Additionally, the Elicitation Study method [160] was employed when validating the outcomes of RQ2 and RQ3; semi-formal interviews with domain experts were conducted to validate the identified interconnections of the navigational CPSs of the C-ES and the proposed security requirements and controls.

3 Overview of the research papers

This section provides an overview of the research papers that constitute this thesis. Paper I addresses Research Question 1 and the subquestions therein. It proposed a reference architecture for the C-ES, and used it to analyze variants of the C-ES. Paper II addresses Research Question 2.1 and Research Question 3. It proposed modifications to the STRIDE and DREAD methods so as to make them applicable to CPSs; it employed these methods to analyze the security risks of CPSs on board the C-ES; and it proposed a systematic approach to identify appropriate security baseline controls to mitigate such risks. Paper III addresses Research Question 2.2, focusing on security requirements alone. It proposed a process for eliciting security requirements for CPSs, building upon the Secure Tropos methodology, and it applied it to the C-ES ecosystem, focusing on the three most vulnerable CPSs of the C-ES. Papers IV and V also address Research Question 2.2, focusing on the joint elicitation of safety and security requirements for CPSs. Specifically, Paper IV reviewed existing security and safety co-engineering approaches, and proposed a multi-attribute taxonomy to analyze them. The analysis highlighted the need for developing a method for the joint elicitation of security and safety requirements for CPSs still at the design stage. Paper V proposed such a method, called SafeSecTropos, and used it to identify safety and security requirements of the three most vulnerable CPSs of the C-ES. Paper VI addresses Research Question 2.3. It proposed a systematic method for discovering and analyzing attack paths in real-world scale interconnected Cyber Physical Systems, and applied it to the navigational CPSs of the C-ES. Paper VII addresses Research Question 2.4. It proposed a reference architecture for the cyber-physical ranges, and used it to describe the structure of a testbed for testing the security posture of the navigation systems of the C-ES. Finally, Paper VIII addresses Research Question 3.1. It proposed a method for assessing the aggregate security risk of large scale, complex CPSs comprising interconnected and interdependent components, by using risk measures of its individual components and the information and control flows among these components. It also proposed a method for selecting a set of effective and efficient security controls among those in an established knowledge base, that reduce the aggregate residual risk whilst minimizing the cost. Both methods were applied to the navigational systems of the remotely controlled ship and the autonomous ship. The mapping between the articles and the research questions that each one addressed is illustrated in Table 4.

Articles	Research Questions
Paper I	RQ1
Paper II	RQ2, RQ 2.1, & RQ 3
Paper III, IV, & V	RQ 2.2
Paper VI	RQ 2.3
Paper VII	RQ 2.4
Paper VIII	RQ 3

Table 4: Articles and Research Questions

3.1 Paper I: Modelling Shipping 4.0: A Reference Architecture for the Cyber-Enabled Ship

A Reference Architecture describes the structure of a system, with its element types and their structures, as well as their interaction types, among each other and with their environment. By describing these, a Reference Architecture defines restrictions for an instantiation (concrete architecture). Through abstraction from individual details, a Reference Architecture is universally valid within a specific domain. Further architectures with the same functional requirements can be constructed based on the reference architecture. Reference architectures, by modeling the system at a level of abstraction free from details of individual instances, facilitate the study of systems that are still in their early stages of development, such as the C-ES.

In this paper we extended the MAF [50] to allow the representation of contemporary systems and technologies in the maritime domain, including the C-ESs, and we developed the extended e-MAF. We then used the e-MAF to develop descriptions of the architecture of vessels with varying level of autonomy, and to identify their functional and operational requirements. Additionally, we identified the interdependencies and interconnections among the CPSs that are components of the C-ES, and we developed graph representations of these. By employing graph analysis metrics we analyzed the criticality of each CPS by considering their trustworthiness, the percentage of the paths that pass through each system, and the number of the connections that each CPS has. The analysis focused on both connections/interconnections and dependencies/interdependencies of the systems.

The outcome of this research provides a comprehensive picture of the C-ES's ecosystem. In particular, these results constitute a stepping stone towards systematically describing the architecture of C-ESs in a harmonized manner; towards assessing and managing the security risks of the C-ES; and towards eliciting the security requirements of the C-ES ecosystem.

3.2 Paper II: Managing Cyber Security Risks of the Cyber-Enabled Ship

Security risk is associated with the potential that threats will exploit vulnerabilities of an asset or group of assets and thereby cause harm. It is assessed in terms of its elements, namely the likelihood of a threat occurring, the extent of the vulnerabilities to the threat, and the magnitude of the impact.

The STRIDE [122] and DREAD [123] methods can effectively analyze security risks in highly interconnected CPSs comprising heterogeneous components, and they are most appropriate for analyzing systems under development, whose operational and functional requirements are not established yet. Alternative approaches need such requirements to produce valid results. In contrast, STRIDE and DREAD facilitate the analysis of conceptual systems by answering questions regarding the security objectives of the targeted ecosystem. Because STRIDE is quantitative and DREAD is qualitative, when used together they provide a holistic view of cyber risk, which cannot be captured by other methods. Further, this hybrid approach facilitates the communication of the results to relevant stakeholders, while allowing the representation of cyber risk in numeric form, thus facilitating the assessment of the effectiveness of security controls at later stages of the risk treatment process.

In Paper II we conceptually modified STRIDE and DREAD to enable the capture of aspects of CPSs on board the C-ES. These modifications pertain to the likelihood and impact criteria of STRIDE and to the criteria used for estimating security risks with DREAD. We also proposed a systematic approach for selecting appropriate cyber security baseline controls to mitigate the estimated risks among those listed in the Industrial Control Systems (ICS) overlay of the NIST Guide to ICS Security. We used the modified STRIDE and DREAD methods to assess the security risks of CPSs on board the C-ES, and we then applied our proposed systematic approach for security control selection to the three most vulnerable on-board systems of the C-ES (AIS, ECDIS, GMDSS). The results support and inform the design of a security architecture for the C-ES.

3.3 Paper III: Shipping 4.0: Security requirements for the Cyber-Enabled Ship

The risk management process is informed by the security requirements. The initial selection of the security controls and techniques is based on systematically analyzed security requirements that the accordant security controls need to satisfy.

In Paper III we proposed a security requirements elicitation process for the C-ESs, building upon the Secure Tropos methodology. The proposed process consists of three stages. The first stage analyzes the ecosystem's actors, goals, assets, and resources, and results in developing an initial actor diagram. The second stage describes the system under study along with the functional and non functional requirements of the CPSs under analysis. The security analysis is performed in the third stage, where the security constraints are identified. We identified the environmental constraints of the C-ES's ecosystem, by employing the e-MAF, as proposed in Paper I. Additionally, we analyzed the actors, goals, processes, and plans of the ecosystem, taking into account the aforementioned constraints.

The application of the proposed process to the C-ES's ecosystem, and in particular to the three most vulnerable CPSs (AIS, ECDIS, and GMDSS) produced a set of common security requirements and a set of system specific security requirements. These were categorized following the classification scheme of the ISO 27001:2013 and ISO 27002:2013 standards. The outcome of this research informs the joint elicitation of security and safety requirements; this is addressed in Paper V.

3.4 Paper IV: Cybersecurity and safety co-engineering of cyber-physical systems: A comprehensive survey

CPSs are characterized by strong coupling between physical and cyber components; therefore, a cyber attack may result in harm affecting both safety and security attributes. This is because there exist strong dependencies between the two domains, even though cases where they are independent do also exist. Three types of such dependencies have been identified, namely conditional dependencies; reinforcement; and conflict [162]. Accordingly, security and safety co-engineering approaches have emerged. These are classified into security-informed safety approaches, i.e. approaches that extend the scope of safety engineering by adapting security-related techniques; safety-informed security approaches, i.e. approaches that extend the scope of security engineering by adapting safety-related techniques; and combined safety and security approaches [163].

In Paper IV we revisited previous surveys on security and safety co-engineering approaches; we reported on the results of a systematic literature survey of such approaches that had not been reviewed before; we proposed a multi-attribute taxonomy of such approaches; and we used it to analyze them. Further, we discussed pertinent open issues and research challenges. The outcome is a comprehensive discussion on the recent advances in security and safety co-engineering, and on open issues, not fully addressed by existing approaches for security and safety co-engineering, that give rise to research challenges. In particular, the need for a holistic, integrated, graphical model based, safety and security requirements elicitation co-engineering approach, applicable to interconnected CPSs, and for systems still at the design stage, was identified.

3.5 Paper V: SafeSec Tropos: Joint security and safety requirements elicitation

The maritime domain is highly dependent on the safety standards and regulations developed by ISO/TC 8/SC 1 and the IMO. Additionally, the increasing digitilization of the domain created the need for security standards and regulations. To this end, the IMO prepared the ISM code, supported by the IMO Resolution MSC.428(98), that incorporated security principles and recommendations into the existing safety risk management process [164]. This in turn calls for the joint elicitation of security and safety requirements.

In Paper V we proposed SafeSec Tropos, a novel integrated method for safety and security requirements engineering for CPSs at the design stage of the system lifecycle. SafeSec Tropos is based on the Secure Tropos method and the STPA that originate from the security and the safety domain respectively. SafeSec Tropos identifies security and safety objectives; it systematically elicits a comprehensive list of requirements; and it links these requirements to objectives, thus facilitating the process of resolving conflicts between security and safety reuirements. An important characteristic of SafeSec Tropos is that it models the system for both safety and security purposes under the same model and provides documentation regarding the potential conflicts of the identified requirements. These conflicts can be resolved by tracing them back to the corresponding safety and security objectives. Further, complex

systems can be analyzed by leveraging the modeling language of the Secure Tropos and the system perspective of the STPA. We applied SafeSec Tropos to the most vulnerable CPSs on board the C-ES, namely the AIS, the ECDIS and the GMDSS.

The outcome of this research was the definition of the safety and security objectives of these systems, and the identification of their safety and security requirements. Such requirements are used as an input in the systematic process described in Paper II and Paper VIII to select the baseline and the optimal security controls for the C-ES.

3.6 Paper VI: Attack Path Analysis for Cyber Physical Systems

The identification and the analysis of potential paths that an adversary may exploit to attack Cyber Physical Systems comprising subsystems enables the comprehensive understanding of the attacks and the impact that they may have to the overall system. This in turn facilitates the definition of appropriate security controls that will satisfy the pertinent security requirements. In a system of networked assets, whereby an asset may well be a system in its own right, an attack path is an ordered sequence of assets that can be used as stepping stones by an attacker seeking to attack one or more assets on the path. The analysis of the attack path is usually based on the vulnerabilities of the systems on the path. This limits considerably the insight into the possible attack and the selection of appropriate security controls, since the focus is on controls that mitigate only the system's vulnerabilities, tending to neglect controls that reduce the likelihood of the threats and the extent of the impact, as well as their combination.

In Paper VI we proposed a novel systematic method for discovering and analyzing attack paths in real-world scale interconnected Cyber Physical Systems. The proposed method aims to discover and analyze attack paths between selected entry and target CPSs, by considering the system's criticality and the overall cyber risk. Compared to existing alternatives, the method handles the scalability problem of attack graphs by considering highly critical nodes and analyzes the resulting paths by considering the cyber risk that each of these represents to the overall system rather than only considering vulnerabilities. We modeled the CPS as a directed graph G(V,E) whose nodes represent the sub-systems and whose edges represent interconnections between nodes, and we built upon results of earlier works to assess the criticality of each system by employing novel graph metrics; the security risks of each CPS that were assessed in Paper II; and by integrating the stakeholders' views by means of a metric that captures the impact of the failure of a component as seen from the stakeholers' perspective. We then applied the proposed method to the navigational CPSs of the C-ES. We identified five critical systems, and five critical attack paths. By leveraging the aforementioned results, appropriate security controls can be defined that will alter the possible attack paths and decrease the risk.

3.7 Paper VII: Towards a cyber-physical range

The assessment of the security posture of CPSs, as well as the evaluation of the effectiveness and efficiency of the accordant controls by means of testing are of paramount importance.

Assessing the security posture in real world CPSs is not recommended, particularly in critical sectors where the continuous, smooth operation of such systems is of vital importance. Hence, such testing is usually performed in a specially configured experimentation platform. Such platforms are known as testbeds and facilitate the testing process by providing physical, virtual, and hybrid models of the system under study. Even though a number of testbeds have been developed for studying different cyber-physical systems, only few have been designed to allow cyber security experimentation; allowing for such functionality is an architectural design issue. A testbed with functionality allowing the testing of the security posture of CPSs constitutes a "cyber-physical range".

In Paper VII we proposed a reference architecture of the cyber-physical ranges. We first surveyed existing CPS testbeds with security testing capacity, we identified common architectural features, and we defined requirements enabling the assessment of the security posture of a system in such testbeds. The proposed reference architecture consists of four modules, namely the control center, the physical components, the virtual components, and the security defensive mechanism modules, and it can be instantiated to different domains.

We then used the proposed reference architecture to describe the structure of a testbed for testing the security posture of the navigation systems of the C-ES. The outcome of this research can be used to drive the development of a cyber-physical range for the C-ES and for other domains as well.

3.8 Paper VIII: Cyber risk propagation and optimal selection of cybersecurity controls for complex cyberphysical systems

The C-ES is a large scale CPS that comprises a number of other, interconnected CPSs. This interconnection increases the cyber risk of the overall system, as such risk propagates between and aggregates at component systems. The complexity of the resulting systems-of-systems in many cases results in difficulties in analyzing the cyber risks of the overall system. In such cases, a method for assessing the cyber risk of a system-of-systems based on information regarding the cyber risks of the component systems is useful.

Further, the selection of security controls that will effectively and efficiently treat the cyber risk is commonly performed manually, or at best with limited automated decision support. Again, selecting the appropriate security controls that, when implemented, will minimize the residual risk whilst minimizing the cost of implementation is a difficult task to perform in large scale, complex CPSs.

In Paper VIII we modeled a complex CPS as a digraph whose nodes represent sub-CPSs and whose edges represent information and control flows among these subsystems. By leveraging this model, we proposed a novel method for assessing the aggregate security risk of large scale, complex CPSs comprising interconnected and interdependent components, by using risk measures of its individual components and the information and control flows among these components. Building upon this method, we proposed a novel method, based on evolutionary programming, for selecting a set of effective and efficient security controls among those in an established knowledge base, that reduce the aggregate residual risk whilst minimizing the cost.

We then applied both methods to the navigational systems of the remotely controlled

ship and of the autonomous ship, to select security controls among those listed in the ICS overlay of the NIST Guide to ICS Security. The outcome is the corresponding optimal sets of security controls, that lead to the definition of the security architecture of the corresponding vessels. They have been found to be in line with the results in previous articles that identified the most vulnerable navigational CPSs of the C-ES, and the critical attack paths, whilst also minimizing the global residual risk.

4 Conclusions, Limitations and Future work

This chapter summarizes the contributions of the research described in the thesis; it discusses the limitations; and it outlines possible directions for future research.

4.1 Contributions

This research contributed to furthering the knowledge on CPS security in general and on the security of the C-ES in particular, as follows:

- We proposed a reference architectural model of the C-ES, based on the extended e-MAF that we developed to describe aspects of the C-ES ecosystem, and the CPSs within this ecosystem that we identified, along with their interconnections, dependencies, and interdependencies. The identified CPSs were categorized according to their operational functionality into groups, namely the Bridge, Engine, Shore Control Center, ICT Infrastructure, and Link systems.
- We adapted the STRIDE and DREAD methods for use in CPSs, and applied them to the reference architectural model of the C-ES, to identify potential threats and to assess the accordant risks for each CPS in the C-ES. Three of these CPSs have been found to be the most vulnerable, namely the AIS, ECDIS, and the GMDSS.
- We leveraged the threat analysis results and the reference architectural model to identify the C-ES security requirements. To this end, we proposed and applied a method for identifying such requirements based on the Secure Tropos methodology. Because safety and security challenges co-exist in contemporary CPSs, we reviewed a number of methods for security and safety co-engineering, we proposed a multi attribute taxonomy of such methods, and we identified under-researched issues and challenges. One of these was the need to develop a holistic, integrated, graphical model-based, safety and security requirements elicitation co-engineering approach. We addressed this challenge by developing the SafeSec Tropos approach for jointly analyzing security and safety requirements of CPSs, and we applied it to the C-ES case.
- We proposed a method for cyber attack path discovery and prioritization for CPSs that comprise a number of sub-systems, and we applied the method to identify possible attack paths between the navigational subsystems of the C-ES.

- Security risks propagate in a system comprising other systems; hence, managing the security risks of the overall system requires insight into this process of risk propagation. We proposed a method for analyzing risk propagation and aggregation in complex CPSs utilizing the results of risk assessments of their individual constituents and we applied it to the C-ES case.
- We proposed a systematic approach that uses a set of criteria that take into account the security requirements; the cyber risks; the possible attacks; and the possibly already existing controls, to select appropriate security baseline controls to mitigate such risks for individual CPSs. Additionally, we proposed a method employing evolutionary programming for automating the selection of an optimal set of security controls out of a list of available controls, that will minimize the residual risk and the cost associated with the implementation of these measures. We applied the risk aggregation and control selection methods to the C-ES case to determine the baseline controls in the security architecture.
- Finally, with an eye towards contributing to the future evaluation of the security of the C-ES, we proposed a reference architecture for the next generation of cyber ranges, namely the cyber-physical ranges, and its instantiation for the case of a testbed for testing the cyber security posture of the navigation systems of the C-ES.

4.2 Limitations

The first difficulty that we encountered when this research started was to obtain information on the CPSs comprising the C-ES which would be sufficiently detailed to allow the definition of an architectural structure of the C-ES. Perhaps not surprisingly, the academic literature on the subject is not extensive, and publicly available sources are equally scarce. We overcame this difficulty by combining all the available relevant sources to define such a structure, which we then validated by means of informal interviews with a limited number of domain experts.

Our research resulted in proposing a number of methods for analyzing the security risks and requirements of CPSs, that lead to the design of the security architecture of the CPSs in the C-ES. In order to validate these methods, we would ideally apply them to an entire, real-world design of a C-ES's CPS infrastructure. Unfortunately, we were unable to do so, as such a design was not available, even less so formal models of the infrastructure. The absence of such models prevented us from developing realistic simulator models. In order to overcome these difficulties, we demonstrated the use of the proposed methods by means of case studies.

We have also proposed security controls to mitigate the identified risks and to satisfy the defined security requirements. The satisfaction of the latter was possible to verify by inspection. Their validation would require input from domain experts. These were difficult to get by and engage, as the C-ES is a relatively young development, and expertise on its security requirements and controls and on how these may affect operations is very scarce. We were, eventually able to recruit a limited number of domain experts, from both industry and academia, who were interviewed and validated both the requirements and the security controls. Still, further validation of these results is desirable.

During the course of this research, a number of software tools were built to support the developed methods and approaches. However, these have not been integrated into a toolbox.

4.3 Future work

A number of possible paths for future work are envisaged, some of which are already ongoing. These are outlined as follows:

- This research resulted in proposing a number of methods for analyzing the security of CPSs, and has demonstrated their use in the case of the C-ES. A research task that will apply these results to other CPS domains can be promptly undertaken. This will also allow making comparisons of the results of this thesis with those of other research works.
- A task that can be undertaken when a real-world design of an autonomous or remotely controlled ship becomes available is the use of the proposed approach for analyzing its security requirements and for designing its security architecture. This will allow the thorough evaluation and refinement of the proposed approach as a whole, and of its constituent methods and techniques at a real-world scale.
- We intend to develop a software toolbox that will implement the proposed methods, and to use it to experientially examine the usability of the proposed approach with domain experts and stakeholders, in the C-ES and other critical application domains.
- Our future work plans include the refinement of the cyber-physical range reference model architecture; its instantiation to specific domain environments, including one for a remotely controlled vessel and the associated shore control center; the design and development of a modular cyber-physical range, and its use for experimentation and validation of its effectiveness, efficiency, configurability, and performance.

References

[1] G. Kavallieratos, S. Katsikas, and V. Gkioulos. Modelling shipping 4.0: A reference architecture for the cyber-enabled ship. In *Proceedings of the Asian Conference on Intelligent Information and Database Systems*, pages 202–217. Springer, 2020.

[2] G. Kavallieratos and S. Katsikas. Managing cyber security risks of the cyber-enabled ship. *Journal of Marine Science and Engineering*, 8(10):768, 2020.

[3] G. Kavallieratos, V. Diamantopoulou, and S. Katsikas. Shipping 4.0: Security requirements for the cyber-enabled ship. *IEEE Transactions on Industrial Informatics*, 2020.

[4] G. Kavallieratos, S. Katsikas, and V. Gkioulos. Cybersecurity and safety co-engineering of cyberphysical systems—a comprehensive survey. *Future Internet*, 12(4):65, 2020.

[5] G. Kavallieratos, S. Katsikas, and V. Gkioulos. Safesec tropos: Joint security and safety requirements elicitation. *Computer Standards & Interfaces*, 70:103429, 2020.

[6] G. Kavallieratos and S. Katsikas. Attack path analysis for cyber physical systems. In *Proceedings of the CyberICPS 2020. Lecture Notes in Computer Science book series (LNCS, volume 12501)*, pages 19–33. Springer, 2020.

[7] G. Kavallieratos, S. K Katsikas, and V. Gkioulos. Towards a cyber-physical range. In *Proceedings of the 5th on Cyber-Physical System Security Workshop*, pages 25–34, 2019.

[8] G. Kavallieratos, G. Spathoulas, and S. Katsikas. Cyber risk propagation and optimal selection of cybersecurity controls for complex cyberphysical systems. *Sensors*, 21(5):1691, 2021.

[9] C. Alcaraz, G. Bernieri, F. Pascucci, J. Lopez, and R. Setola. Covert channels-based stealth attacks in industry 4.0. *IEEE Systems Journal*, 13(4):3980–3988, 2019.

[10] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin. Smart factory of industry 4.0: Key technologies, application case, and challenges. *Ieee Access*, 6:6505–6519, 2017.

[11] SINTEF. Shipping 4.0 presented at singapore maritime week. `https://www.sintef.no/en/latest-news/shipping-4.0-presented-at-singapore-maritime-week/`. [Online; accessed 10-02-2020].

[12] C. Alcaraz. Secure interconnection of IT-OT networks in industry 4.0. In *Critical Infrastructure Security and Resilience*, pages 201–217. Springer, 2019.

[13] M. Postráneckỳ and M. Svítek. Smart city near to 4.0—an adoption of industry 4.0 conceptual model. In *2017 Smart City Symposium Prague (SCSP)*, pages 1–5. IEEE, 2017.

[14] A. Mehdiabadi, M. Tabatabeinasab, C. Spulbar, A. Karbassi Yazdi, and R. Birau. Are we ready for the challenge of banks 4.0? designing a roadmap for banking systems in industry 4.0. *International Journal of Financial Studies*, 8(2):32, 2020.

[15] J. Cross and G. Meadow. Autonomous ships 101. *Journal of Ocean Technology*, 12(3):23–27, 2017.

[16] V. Bertram. Technologies for low-crew/no-crew ships. In *Proceedings of the Forum Captain Computer IV*, page 24. Citeseer, 2002.

[17] R. Schönknecht. *Schiffe und Schiffahrt von morgen*. VEB Verlag Technik Berlin, 1973.

[18] MUNIN. Maritime unmanned navigation through intelligence in networks. `http://www.unmanned-ship.org/munin/`, 2016. [Online; accessed 25-06-2020].

[19] DNV-GL. The ReVolt, A new inspirational ship concept. `https://www.dnvgl.com/technology-innovation/revolt/index.html`. [Online; accessed 27-10-2020].

[20] Rolls-Royce. Remote and autonomous ship-the next steps. page 88, 2016.

[21] A. Felski and K. Zwolak. The ocean-going autonomous ship—challenges and threats. *Journal of Marine Science and Engineering*, 8(1):41, 2020.

[22] Yara. Yara birkeland status. `https://www.yara.com/news-and-media/press-kits/yara-birkeland-press-kit/`. [Online; accessed 19-10-2020].

[23] IBM. Mayflower autonomous ship launches. `https://newsroom.ibm.com/2020-09-15-Mayflower-Autonomous-Ship-Launches`. [Online; accessed 19-10-2020].

[24] Norwegian University of Science and Technology. Autoferry - autonomous all-electric passenger ferries for urban water transport. `https://www.ntnu.edu/autoferry`. [Online; accessed 19-10-2020].

[25] K. Tam and K. Jones. MaCRA: a model-based framework for maritime cyber-risk assessment. *WMU Journal of Maritime Affairs*, 18(1):129–163, 2019.

[26] European Network and Information Security Agency - ENISA. Port cybersecurity good practices for cybersecurity in the maritime sector. Technical report, 2019.

[27] K. D. Jones, K. Tam, and M. Papadaki. Threats and impacts in maritime cyber security. *Engineering Technology Reference 1(1)*, 2016.

[28] C. H. Chang, S. Wenming, Z. Wei, P. Changki, and C. A. Kontovas. Evaluating cybersecurity risks in the maritime industry: a literature review. In *Proceedings of the International Association of Maritime Universities (IAMU) Conference*, 2019.

[29] M. S. K. Awan and M. A. Al Ghamdi. Understanding the vulnerabilities in digital components of an integrated bridge system (IBS). *Journal of Marine Science and Engineering*, 7(10):350, 2019.

[30] J Saul. Cyber Threats prompt return of radio for ship navigation. Reuters. `https://www.reuters.com/article/us-shipping-gps-cyber-idUSKBN1AN0HT`, 2017. [Online; accessed 04-05-2021].

[31] BBC News. San Diego port hit by ransomware attack. `https://www.bbc.com/news/technology-45677511`, 2018. [Online; accessed 04-05-2021].

[32] Information Security News. Hacking Attack in port of Barcelona. `https://www.securitynewspaper.com/2018/09/26/hacking-attack-in-port-of-barcelona/`, 2018. [Online; accessed 04-05-2021].

[33] CyberKeel. Maritime cyber-risks. Technical report, 2014.

[34] S. N. Sirimanne, J. Hoffman, W. Juan, R. Asariotis, M. Assaf, G. Ayala, H. Benamara, D. Chantrel, J. Hoffmann, A. Premti, et al. Review of maritime transport, 2019. Technical report, United Nations, 2019.

[35] DNV-GL. Autonomous and remotely operated ships, Class Guideline. Technical report, pp 111, 2018.

[36] BIMCO, CLIA, ICS, INTERCARGO, INTERTANKO, OCIMF and IUMI. The Guidelines on Cyber Security Onboard Ships, Version 4, pp 51. Technical report, 2020.

[37] SINTEF, TCOMS. The research institutes' roadmap towards smart and autonomous maritime transport systems. Technical report, 2020.

[38] G. Reilly and J. Jorgensen. Classification considerations for cyber safety and security in the smart ship era. In *Proceedings of the International Smart Ships Technology Conference*, pages 26–27, 2016.

[39] International Maritime Organization. IMO takes first steps to address autonomous ships. `http://www.imo.org/en/mediacentre/pressbriefings/pages/08-msc-99-mass-scoping.aspx`, 2018. [Online; accessed 19-10-2020].

[40] Lloyds Register. Cyber-enabled ships: Deploying information and communications technology in shipping–lloyds register's approach to assurance. Technical report, 2016.

[41] Lloyd's Register. Ship right procedure assignment for cyber descriptive notes for autonomous remote access ships. Technical report, 2017.

[42] DNV-GL. Remote controlled and autonomous ships. Technical report, 2018.

[43] China Classification Society. Guidelines for autonomous cargo ships. Technical report, 2018.

[44] C. Alcaraz and J. Lopez. Analysis of requirements for critical control systems. *International journal of critical infrastructure protection*, 5(3-4):137–145, 2012.

[45] M. Al-Kuwaiti, N. Kyriakopoulos, and S. Hussein. A comparative analysis of network dependability, fault-tolerance, reliability, security, and survivability. *IEEE Communications Surveys & Tutorials*, 11(2):106–124, 2009.

[46] J. C. Knight and E. A. Strunk. Achieving critical system survivability through software architectures, architecting dependable systems. In *Architecting Dependable Systems II. LNCS 3069*. Citeseer, 2004.

[47] J. Bowen and V. Stavridou. Safety-critical systems, formal methods and standards. *Software engineering journal*, 8(4):189–209, 1993.

[48] F. Skopik and P. Smith. *Smart grid security: Innovative solutions for a modernized grid*. Syngress, 2015.

[49] International Organization for Standardization (ISO). Internet of Things (IoT) — Reference Architecture, NEK ISO 30141:2018. Technical report, 2018.

[50] H.C. Mayr and M. Pinzger. A domain-specific architecture framework for the maritime domain. *INFORMATIK*, 2016:773–784, 2016.

[51] Reference Architecture Model Industrie 4.0 (RAMI 4.0). Standard DIN SPEC 91345:2016-04, DIN Deutsches Institut für Normung e. V., Berlin, Germany, 2016.

[52] J. Durand G. Bleakley A. Chigani R. Martin B. Murphy S. W. Lin, B. Miller and M. Crawford. The Industrial Internet of Things Volume G1: Reference Architecture. Technical report, Industrial Internet Consortium, 2019.

[53] M. Moghaddam, M. N. Cadavid, C. R. Kenley, and A. V. Deshmukh. Reference architectures for smart manufacturing: A critical review. *Journal of manufacturing systems*, 49:215–225, 2018.

[54] B. Di Martino, M. Rak, M. Ficco, A. Esposito, S.A. Maisto, and S. Nacchia. Internet of things reference architectures, security and interoperability: A survey. *Internet of Things*, 1-2:99–112, 2018.

[55] J. Augusto-Gonzalez, A. Collen, S. Evangelatos, M. Anagnostopoulos, G. Spathoulas, K. M Giannoutakis, K. Votis, D. Tzovaras, B. Genge, E. Gelenbe, et al. From internet of threats to internet of things: A cyber security architecture for smart homes. In *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pages 1–6. IEEE, 2019.

[56] A. Cox, P. Parslow, B. De Lathouwer, E. Klien, B. Kempen, and J. Lonien. D4. 2: Definition of Smart City Reference Architecture. *ESPRESSO systEmic Standardisation apPRoach to Empower Smart citieS and cOmmunities*, 2016.

[57] I. Schieferdecker, N. Tcholtchev, P. Lämmel, R. Scholz, and E. Lapi. Towards an open data based ICT reference architecture for smart cities. In *2017 Conference for E-Democracy and Open Government (CeDEM)*, pages 184–193. IEEE, 2017.

[58] CEN-CENELEC-ETSI Smart Grid Coordination Group. Smart Grid Reference Architecture. Technical report, 2012.

[59] A. Rashid, H. Chivers, G. Danezis, E. Lupu, and A. Martin. The cyber security body of knowledge. `https://www.cybok.org/media/downloads/cybok_version_1.0.pdf`, 2019. [Online; accessed 27-02-2021].

[60] The Smart Grid Interoperability Panel – Smart Grid Cybersecurity Committee. Guidelines for smart grid cybersecurity: Volume 1 - smart grid cybersecurity strategy, architecture, and high-level requirements. Technical report, 2014.

[61] A technical specification for the common shore-based system architecture (cssa). Technical report, International Association of Marine Aids to Navigation and Lighthouse Authorities, 2015.

[62] Homepage — e-Navigation. `http://www.imo.org/en/OurWork/Safety/Navigation/Pages/eNavigation.aspx`. [Online; accessed 10-10-2019].

[63] Ø. J. Rødseth and A . Tjora. A system architecture for an unmanned ship. In *Proceedings of the 13th International Conference on Computer and IT Applications in the Maritime Industries (COMPIT 2014), Redworth, UK*, 2014.

[64] Ø. J. Rødseth, B. Kvamstad, T. Porathe, and H. C. Burmeister. Communication architecture for an unmanned merchant ship. In *Proceedings of the 2013 MTS/IEEE OCEANS - Bergen*, pages 1–9, June 2013.

[65] H.-C. Burmeister, W. Bruhn, L. Walther, J. A. Moræus, and B. Sage-Fuller. MUNIN D8.6: Final Report: Autonomous Bridge. Technical report, 2015.

[66] S. N. MacKinnon, Y. Man, and M. Baldauf. MUNIN D8.8 Final Report Shore Control Centre. Technical report, 2015.

[67] M. Schmidt, E. Fentzahn, G. F. Atlason, and H. Rødseth. MUNIN 8.7 Final Report Autonomous Engine Room. Technical report, 2015.

[68] J. M. Varas, S. Hirdaris, R. Smith, P. Scialla, W. Caharija, Z. Bhuiyan, T. Mills, W. Naeem, L. Hu, I. Renton, et al. MAXCMAS project: Autonomous COLREGs compliant ship navigation. In *Proceedings of the 16th Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT) 2017*, pages 454–464, 2017.

[69] V. Bolbot, G. Theotokatos, E. Boulougouris, and D. Vassalos. A novel cyber-risk assessment method for ship systems. *Safety Science*, 131:104908, 2020.

[70] V. Bolbot, G. Theotokatos, E. Boulougouris, and D. Vassalos. Safety related cyber-attacks identification and assessment for autonomous inland ships. In *Proceedings of the International Seminar on Safety and Security of Autonomous Vessels (ISSAV)*, 2019.

[71] M. Chaal, O. V. Banda, S. Basnet, S. Hirdaris, and P. Kujala. An initial hierarchical systems structure for systemic hazard analysis of autonomous ships. In *Proceedings of the International Seminar on Safety and Security of Autonomous Vessels (ISSAV) and European STAMP Workshop and Conference (ESWC) 2019*, pages 140–153. Sciendo, 2020.

[72] K. Heffner and Ø. J. Rødseth. Enabling technologies for maritime autonomous surface ships. In *Proceedings Journal of Physics: Conference Series*, volume 1357, page 012021. 2019.

[73] I. Im, D. Shin, and J. Jeong. Components for smart autonomous ship architecture based on intelligent information technology. *Procedia computer science*, 134:91–98, 2018.

[74] M. Höyhtyä, J. Huusko, M. Kiviranta, K. Solberg, and J. Rokka. Connectivity for autonomous ships: Architecture, use cases, and research challenges. In *Proceedings of the 2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 345–350. IEEE, 2017.

[75] J. Poikonen. Requirements and challenges of multimedia processing and broadband connectivity in remote and autonomous vessels. In *Proceedings of the 2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 1–5. IEEE, 2018.

[76] International Maritime Organization. Guidelines on maritime cyber risk management. Technical report, 5 July 2017.

[77] American Bureau of Shipping. Guidance notes on the application of cybersecurity principles to marine and offshore operations abs cybersafetytm volume 1. Technical report, 2016.

[78] DNV-GL. Cyber security resilience management for ships and mobile offshore units in operation. Technical report, 2016.

[79] H. Boyes and R. Isbell. Code of practice cyber security for ships. Technical report, The Institution of Engineering and Technology, 2017.

[80] K. Gevers, L. Oelschläger, J. Schirrmacher, et al. IT-Grundschutz Profile for Shipping Companies Minimum Protection for Ship Operations. Technical report, 2020.

[81] J. Kouns and D. Minoli. *Information technology risk management in enterprise environments*. John Wiley Sons, Inc., 2010.

[82] G. Wangen, C. Hallstensen, and E. Snekkenes. A security risk analysis model for information systems: Causal relationships of risk factors and vulnerability propagation analysis. *International Journal of Information Security*, 17:681 – 699., 2018.

[83] European Network and Information Security Agency (ENISA). Inventory of risk assessment and risk management methods. Technical report, 2006.

[84] P.L. Campbell and J.E. Stamp. A Classification Scheme for Risk Assessment Methods. Technical report, 2004.

[85] S. Ali, T. Al Balushi, Z. Nadir, and O.K. Hussain. Risk Management for CPS Security. In *Proceedings of the Cyber Security for Cyber Physical Systems*, pages 11 – 34. Springer International Publishing AG, Cham, Switzerland, 2018.

[86] X. Lyu, Y. Ding, and S.-H. Yang. Safety and security risk assessment in cyber-physical systems. *IET Cyber-Physical Systems: Theory & Applications*, 2019.

[87] V. Lamba, N. Šimková, and B. Rossi. Recommendations for smart grid security risk management. *Cyber-Physical Systems*, 5(2):92 – 118, 2019.

[88] K. Kandasamy, S. Srinivas, K. Achuthan, and V.P. Rangan. IoT cyber risk: a holistic analysis of cyber risk assessment frameworks, risk vectors, and risk ranking process. *EURASIP Journal on Information Security*, 8:18, 2020.

[89] Y. Cherdantseva, P. Burnap, A. Blyth, P. Eden, K. Jones, H. Soulsby, and K. Stoddart. A review of cyber security risk assessment methods for scada systems. *Computers & security*, 56:1–27, 2016.

[90] G. Macher, E. Armengaud, E. Brenner, and C. Kreiner. Threat and Risk Assessment Methodologies in the Automotive Domain. *Procedia Computer Science*, 83:1288–1294, 2016.

[91] J. E. Vinnem and I. B. Utne. Risk from cyberattacks on autonomous ships. *Safety and Reliability-Safe Societies in a Changing World*, 2018.

[92] K. Tam and K. Jones. Cyber-risk assessment for autonomous ships. In *Proceedings of the 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pages 1–8. IEEE, 2018.

[93] I. S. Shipunov, K. S. Voevodskiy, A. P. Nyrkov, Y. F. Katorin, and Y. A. Gatchin. About the problems of ensuring information security on unmanned ships. In *Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pages 339–343. IEEE, 2019.

[94] K. Bernsmed, C. Frøystad, P. H. Meland, D. A. Nesheim, and Ø. J. Rødseth. Visualizing cyber security risks with bow-tie diagrams. In *Proceedings of the International Workshop on Graphical Models for Security*, pages 38–56. Springer, 2017.

[95] B. Silverajan, M. Ocak, and B. Nagel. Cybersecurity attacks and defences for unmanned smart ships. In *Proceedings of the 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pages 15–20. IEEE, 2018.

[96] Y. G. Kim, D. Jeong, S. H. Park, J. Lim, and D. K. Baik. Modeling and simulation for security risk propagation in critical information systems. In *Proceedings of the International Conference on Computational and Information Science*, pages 858–868. Springer, 2006.

[97] S. Kondakci. A new assessment and improvement model of risk propagation in information security. *International Journal of Information and Computer Security*, 1(3):341–366, 2007.

[98] N. Feng, H. J. Wang, and M. Li. A security risk analysis model for information systems: Causal relationships of risk factors and vulnerability propagation analysis. *Information sciences*, 256:57–73, 2014.

[99] H. Orojloo and M. A. Azgomi. A method for evaluating the consequence propagation of security attacks in cyber–physical systems. *Future Generation Computer Systems*, 67:57–71, 2017.

[100] T. Wang, X. Wei, T. Huang, J. Wang, L. Valencia-Cabrera, Z. Fan, and M. J. Pérez-Jiménez. Cascading failures analysis considering extreme virus propagation of cyber-physical systems in smart grids. *Complexity*, 2019, 2019.

[101] S. König, S. Rass, S. Schauer, and A. Beck. Risk propagation analysis and visualization using percolation theory. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 7(1), 2016.

[102] Z. Qu, Y. Zhang, N. Qu, L. Wang, Y. Li, and Y. Dong. Method for quantitative estimation of the risk propagation threshold in electric power CPS based on seepage probability. *IEEE Access*, 6:68813–68823, 2018.

[103] A. A. Malik and D. K. Tosh. Quantitative risk modeling and analysis for large-scale cyber-physical systems. In *Proceedings of the 29th International Conference on Computer Communications and Networks (ICCCN)*, pages 1–6. IEEE, 2020.

[104] M. Theoharidou, P. Kotzanikolaou, and D. Gritzalis. A multi-layer criticality assessment methodology based on interdependencies. *Computers & Security*, 29(6):643–658, 2010.

[105] M. Theoharidou, P. Kotzanikolaou, and D. Gritzalis. Risk assessment methodology for interdependent critical infrastructures. *International Journal of Risk Assessment and Management*, 15(2-3):128–148, 2011.

[106] X. Zhou, F. Wang, and Y. Ma. An overview on energy internet. In *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 126–131. 2015.

[107] Q. Hong, T. Jianwei, T. Zheng, Q. Wenhui, L. Chun, L. Xi, and Z. Hongyu. An information security risk assessment algorithm based on risk propagation in energy internet. In *Proceedings of the IEEE Conference on Energy Internet and Energy System Integration (EI2)*, pages 1–6. IEEE, 2017.

[108] S. Li, S. Zhao, Y. Yuan, Q. Sun, and K. Zhang. Dynamic security risk evaluation via hybrid bayesian risk graph in cyber-physical social systems. *IEEE Transactions on Computational Social Systems*, 5(4):1133–1141, 2018.

[109] A. Karbowski and K. Malinowski. Two-level system of on-line risk assessment in the national cyberspace. *IEEE Access*, 8:181404–181410, 2020.

[110] A. Schilling and B. Werners. Optimal selection of it security safeguards from an existing knowledge base. *Eur. J. Oper. Res.*, 248:318–327, 2016.

[111] P. Nespoli, D. Papamartzivanos, F. Gómez Mármol, and G. Kambourakis. Optimal countermeasures selection against cyber attacks: A comprehensive survey on reaction frameworks. *IEEE Commun. Surv. Tutor.*, 20(2):1361–1396, 2018.

[112] S. Hussain, A. Kamal, S. Ahmad, G. Rasool, and S. Iqbal. Threat modelling methodologies: a survey. *Sci. Int.(Lahore)*, 26(4):1607–1609, 2014.

[113] F. Sidi, A. J. Marzanah, L. S. Affendey, I. Ishak, I. Sharef, M. Zolkepli, M. Ming, M. F. Abd Mokhti, M. Daud, N. Zainuddin, et al. A comparative analysis study on information security threat models: a propose for threat factor profiling. *J. Eng. Appl. Sci*, 12(3):548–554, 2017.

[114] N. Shevchenko, T. A. Chick, P. O'Riordan, T. P. Scanlon, and C. Woody. Threat modeling: a summary of available methods. Technical report, Carnegie Mellon University Software Engineering Institute Pittsburgh United, 2018.

[115] Y. Ayrour, A. Raji, and M. Nassar. Modelling cyber-attacks: a survey study. *Network Security*, 2018(3):13–19, 2018.

[116] H. Al-Mohannadi, Q. Mirza, A. Namanya, I. Awan, A. Cullen, and J. Disso. Cyber-attack modeling analysis techniques: An overview. In *Proceedings of the IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pages 69–76. IEEE, 2016.

[117] D. Hanic and A. Surkovic. An attack model of autonomous systems of systems, 2018.

[118] Y. C. Chen, V. Mooney, and S. Grijalva. A survey of attack models for cyber-physical security assessment in electricity grid. In *Proceedings of the IFIP/IEEE 27th International Conference on Very Large Scale Integration (VLSI-SoC)*, pages 242–243. IEEE, 2019.

[119] J. Zeng, S. Wu, Y. Chen, R. Zeng, and C. Wu. Survey of attack graph analysis methods from the perspective of data and knowledge processing. *Security and Communication Networks*, 2019.

[120] I. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz, and J. Lopez. A survey of iot-enabled cyberattacks: Assessing attack paths to critical infrastructures and services. *IEEE Communications Surveys Tutorials*, 20(4):3453–3495, 2018.

[121] G. Kavallieratos, S. Katsikas, and V. Gkioulos. Cyber attacks against the autonomous ship. In *Proceedings of the SECPRE 2018, CyberICPS 2018. Lecture Notes in Computer Science, vol 11387*, pages 20–36. Springer, 2018.

[122] A. Shostack. *Threat Modeling: Designing for Security.* Wiley Publishing, 1st edition, 2014.

[123] Microsoft. Chapter 3 – Threat Modeling, 2010. [online] `https://docs.microsoft.com/en-us/previous-versions/msp-n-p/ff648644(v=pandp.10)?redirectedfrom=MSDN`.

[124] G. Kavallieratos, V. Gkioulos, and S. K. Katsikas. Threat analysis in dynamic environments: The case of the smart home. In *Proceedings of the 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 234–240. IEEE, 2019.

[125] R. Khan, K. McLaughlin, D. Laverty, and S. Sezer. STRIDE-based threat modeling for cyber-physical systems. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6. IEEE, 2017.

[126] Adi Karahasanovic, Pierre Kleberger, and Magnus Almgren. Adapting threat modeling methods for the automotive industry. In *Proceedings of the 15th ESCAR Conference*, pages 1–10, 2017.

[127] A. Nhlabatsi, B. Nuseibeh, and Y. Yu. Security requirements engineering for evolving software systems: A survey. In *Security-aware systems applications and software development methods*, pages 108–128. IGI Global, 2012.

[128] N. Mead. How to compare the security quality requirements engineering (square) method with other methods. Technical report, August 2007.

[129] C. Raspotnig and A. Opdahl. Comparing risk identification techniques for safety and security requirements. *Journal of Systems and Software*, 86(4):1124 – 1151, 2013. SI : Software Engineering in Brazil: Retrospective and Prospective Views.

[130] H. Mouratidis and P. Giorgini. Secure tropos: a security-oriented extension of the tropos methodology. *International Journal of Software Engineering and Knowledge Engineering*, 17(02):285–309, 2007.

[131] P. Bresciani, A. Perini, P. Giorgini, F. Giunchiglia, and J. Mylopoulos. Tropos: An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems*, 8(3):203–236, 2004.

[132] R. Matulevičius, N. Mayer, H. Mouratidis, E. Dubois, P. Heymans, and N. Genon. Adapting secure tropos for security risk management in the early phases of information systems development. In *Proceedings of the International Conference on Advanced Information Systems Engineering*, pages 541–555. 2008.

[133] C. Kalloniatis, V. Diamantopoulou, K. Kotis, C. Lyvas, K. Maliatsos, M. Gay, A. Kanatas, and C. Lambrinoudakis. Towards the design of an assurance framework for increasing security and privacy in connected vehicles. *International Journal of Internet of Things and Cyber-Assurance*, 2019.

[134] H. Mouratidis and V. Diamantopoulou. A security analysis method for industrial internet of things. *IEEE Transactions on Industrial Informatics*, 14(9):4093–4100, Sep. 2018.

[135] D. Muñante, V. Chiprianov, L. Gallon, and P. Aniorté. A review of security requirements engineering methods with respect to risk analysis and model-driven engineering. In *Proceedings of International Conference on Availability, Reliability, and Security*, pages 79–93. Springer, 2014.

[136] D. Mellado, C. Blanco, L. E. Sánchez, and E. Fernández-Medina. A systematic review of security requirements engineering. *Computer Standards & Interfaces*, 32(4):153–165, 2010.

[137] V. Bolbot, G. Theotokatos, L. M. Bujorianu, E. Boulougouris, and D. Vassalos. Vulnerabilities and safety assurance methods in cyber-physical systems: A comprehensive review. *Reliability Engineering & System Safety*, 182:179–193, 2019.

[138] C. Raspotnig and A. Opdahl. Comparing risk identification techniques for safety and security requirements. *Journal of Systems and Software*, 86(4):1124–1151, 2013.

[139] N. Leveson and J. Thomas. *STPA handbook*. 2018.

[140] I. Friedberg, K. McLaughlin, P. Smith, D. Laverty, and S. Sezer. STPA-SafeSec: Safety and security analysis for cyber-physical systems. *Journal of information security and applications*, 34:183–196, 2017.

[141] W. E. Young. STPA-SEC for cyber security mission assurance. *Eng Syst. Div. Syst. Eng. Res. Lab*, 2014.

[142] E. Lisova, I. Sljivo, and A. Causevic. Safety and security co-analyses: A systematic literature review. *IEEE Systems Journal*, 13, 2018.

[143] S. Kriaa, L. Pietre-Cambacedes, M. Bouissou, and Y. Halgand. A survey of approaches combining safety and security for industrial control systems. *Reliability engineering & system safety*, 139:156–178, 2015.

[144] ITEA MERgE Project. `http://www.merge-project.eu/`, 2016. [Online; accessed 25-10-2020].

[145] M. A. Lundteigen and B. A. Gran. The need of improved methods to handle functional safety and cybersecurity in industrial control and safety systems. In *Proceedings of Enlarged Halden Programme Group Meeting*. OECD Halden reaktorprosjektet, 2019.

[146] M. Sun, S. Mohan, L. Sha, and C. Gunter. Addressing safety and security contradictions in cyber-physical systems. In *Proceedings of the 1st Workshop on Future Directions in Cyber-Physical Systems Security (CPSSW'09)*. Citeseer, 2009.

[147] D. Pereira, C. Hirata, R. Pagliares, and S. Nadjm-Tehrani. Towards combined safety and security constraints analysis. In Stefano Tonetta, Erwin Schoitsch, and Friedemann Bitsch, editors, *Computer Safety, Reliability, and Security*, pages 70–80, Cham, 2017. Springer International Publishing.

[148] M. Höyhtyä, J. Huusko, M. Kiviranta, K. Solberg, and J. Rokka. Connectivity for autonomous ships: Architecture, use cases, and research challenges. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 345–350. IEEE, 2017.

[149] Bureau Veritas. Guidelines for autonomous shipping. Technical report, 2017.

[150] DNVGL. Cyber security capabilities of control system components. Technical report, 2018.

[151] F. Asplund, J. McDermid, R. Oates, and J. Roberts. Rapid integration of CPS security and safety. *IEEE Embedded Systems Letters*, 11(4):111–114, 2019.

[152] J. Glomsrud and J. Xie. A Structured STPA Safety and Security Co-analysis Framework for Autonomous Ships. In *Proceedings of the Safety and Reliability–Safe Societies in a Changing World. ESREL 2018, June 17-21, 2018, Trondheim, Norway*. Taylor & Francis, 2019.

[153] N. H. C. Guzman, D. Kwame M. Kufoalor, I. Kozin, and M A. Lundteigen. Combined safety and security risk analysis using the UFoI-E method: A case study of an autonomous surface vessel. In *29th European Safety and Reliability Conference*, pages 4099–4106, 2019.

[154] E. N. Torkildson, J. Li, S. O. Johnsen, and J. A. Glomsrud. Empirical studies of methods for safety and security co-analysis of autonomous boat. *Safety and Reliability–Safe Societies in a Changing World. Proceedings of ESREL 2018, June 17-21, 2018, Trondheim, Norway*, 2018.

[155] L. Kretschmann, Ø. J. Rødseth, B. S. Fuller, H. Noble, J. Horahan, and H. McDowell. MUNIN D9.3: Quantitative assessment. Technical report, 2015.

[156] L. Kretschmann, Ø. J. Rødseth, A. Tjora, B. Sage Fuller, H. Noble, and J. Horahan. MUNIN D9.2: Qualitative assessment. Technical report, 2015.

[157] P. Johannesson and E. Perjons. *An introduction to design science*. Springer, 2014.

[158] A. Dresch, D. Pacheco Lacerda, and Paulo Augusto Cauchick M. A distinctive analysis of case study, action research and design science research. *Revista brasileira de gestão de negócios*, 17(56):1116, 2015.

[159] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77, 2007.

[160] T. W. Edgar and D. O. Manz. *Research methods for cyber security*. Syngress, 2017.

[161] H. Snyder. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104:333–339, 2019.

[162] L. Pièvre-Cambacédès and M. Bouissou. Modeling safety and security interdependencies with bdmp (boolean logic driven markov processes). In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 2852–2861. IEEE, 2010.

[163] S. Paul and L. Rioux. Over 20 years of research into cybersecurity and safety engineering: a short bibliography. In *Proceedings of the Safety and Security Engineering VI*, pages 335–349. 2015.

[164] International Maritime Organization IMO. Maritime cyber risk. https://www.imo.org/en/OurWork/Security/Pages/Cyber-security.aspx. [Online; accessed 16-12-2020].

Part II: Research Articles

5 Article I: Modelling Shipping 4.0: A Reference Architecture for the Cyber-Enabled Ship [1]

Modelling Shipping 4.0: A Reference Architecture for the Cyber-Enabled Ship

12th Asian Conference, ACIIDS 2020, Phuket, Thailand, March 23–26, 2020, Proceedings, Part II, Springer, (LNCS, volume 12034 and LNAI, volume 12034)

Georgios Kavallieratos[1][0000 0003 1278 1943], Sokratis Katsikas[1,2][0000 0003 2966 9683], and Vasileios Gkioulos[1][0000 0001 7304 3835]

[1] Norwegian University of Science and Technology, Department of Information Security and Communications Technology, Gjøvik, Norway
(name.surname)@ntnu.no, sokratis.katsikas@ouc.ac.cy
[2] Open University of Cyprus, School of Pure and Applied Sciences, Latsia, Nicosia, Cyprus

Abstract. There is intense activity of the maritime industry towards making remotely controlled and autonomous ships sail in the near future; this activity constitutes the instantiation of the Industry 4.0 process in the maritime industry. Yet, a reference model of the architecture of such vessels that will facilitate the "shipping 4.0" process has not yet been defined. In this paper we extend the existing Maritime Architectural Framework to allow the description of the cyber-enabled ships (C-ESs), and we demonstrate the use of the extended framework by developing descriptions of the architecture of variants of the Cyber-enabled ship. The results can be used not only to systematically describe the architecture of Cyber-enabled ships in a harmonized manner, but also to identify standardization gaps, and to elicit the cybersecurity requirements of the C-ES ecosystem.

Keywords: Autonomous ships · Reference architecture · Cyber-physical systems.

1 Introduction

Industry 4.0 describes the trend towards increasing automation and connectivity, by leveraging technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), and Big Data Analytics. In the maritime sector, despite the fact that nowadays almost all ships are automated in some way, the shipping industry is coming to alignment with Industry 4.0 with the emergence of autonomous vessels [9]. However, this is not a direct process towards a fully autonomous system, but rather a gradual shift towards the digital transformation of maritime operations both ship- and shore-side [19]. In this "Shipping 4.0" process, the interaction and dynamics between ship/land, ship/authorities and ship/ship are expected to change fundamentally [32].

In modern systems engineering, specifications are created by means of employing some requirements engineering process. Such a process is used for eliciting the information needed to create a solution architecture, and subsequently to implement and operate it. Thus, the system architecture is a key element of the process of implementing and deploying a system according to the specifications. For simple systems, this process can be carried out semi-formally, by direct communication among the different teams involved in the process. However, this approach does not work in the case of complex systems-of-systems, where a large number of engineering teams are responsible for different components and parts of the system, and the knowledge and work is much more fragmented. This situation calls for a formalized and governed process, where communication is done in a formal and knowledge-intensive manner and where standards are needed at a certain point. One part of the solution to this problem is to use a method which has proven to be useful, namely the development of a *Reference Architecture* [33].

A Reference Architecture describes the structure of a system, with its element types and their structures, as well as their interaction types, among each other and with their environment. By describing these, a Reference Architecture defines restrictions for an instantiation (concrete architecture). Through abstraction from individual details, a Reference Architecture is universally valid within a specific domain. Further architectures with the same functional requirements can be constructed based on the reference architecture [17], [3].

A Maritime Architecture Framework (MAF) was proposed in [36], to facilitate the development and adoption of new systems and technologies in the maritime domain. The development process of the MAF followed that of the Smart Grid Architectural Model (SGAM) [6]; accordingly, the MAF has been developed taking into consideration existing maritime architectures, including the Common Shore Based System Architecture [4] and the International Maritime Organization's (IMO) e-Navigation architecture [2].

The IMO uses the term *MASS (Maritime Autonomous Surface Ship)* for the autonomous ship [14]. Cyber-Enabled ships (C-ES) are ships that integrate Cyber Physical Systems (CPSs) within their architectures, and whose operations may be fully or partially carried out autonomously. Thus, a C-ES may be a conventional, remotely controlled or autonomous ship. Further, a C-ES can be manned or unmanned, depending on its operational procedures and its infrastructure. According to the IMO, the levels of autonomy for a MASS are defined as follows:

* *AL0: Ship with automated processes and decision support:* Seafarers are on board to operate and control shipboard systems and functions. Some operations may be automated.
* *AL1: Remotely controlled ship (with seafarers on board):* The ship is controlled and operated from another location, but seafarers are on board.
* *AL2: Remotely controlled ship (without seafarers on board):* The ship is controlled and operated from another location. There are no seafarers on board.
* *AL3: Fully autonomous ship:* The operating system of the ship is able to make decisions and determine actions by itself.

AL0 describes the conventional ship, where the C-ES's operations are the same with those of traditional vessels. Although many contemporary ICT systems can be on board in order to support processes related with navigation and engine control, human operators maintain the central role. For the remotely controlled variants (AL1, AL2), most of the ship's systems are capable of performing predefined actions without human intervention. The ship's operations depend on the communication with the Shore Control Center (SCC) and at the same time are influenced from on-board crew and CPSs. The human operator at these levels gives directions and controls the vessel's systems either locally (AL1) or remotely (AL2), whilst operations such as mooring, navigating, cargo loading and unloading are performed entirely by remote control. At the last level of autonomy (AL3), most of the ship's operations rely on the on-board CPSs, although some of the operations may be supervised by a SCC. Furthermore, the ship is equipped with contemporary navigation, engine and control systems, such as collision avoidance systems. At this level, the human vector does not exist and advanced systems are responsible for the availability, maintainability and reliability of the operations.

In this paper we extend the MAF to include autonomous vessels. We then demonstrate the use of the MAF to create architectural instances of autonomous vessels with varying level of autonomy, and we identify their functional and operational requirements. Finally, we identify and analyze the interdependencies and interconnections among the CPSs that are components of the C-ES. The contribution of this work is as follows:

- The development of an extended Maritime Architectural Framework that can accommodate autonomous vessels;
- The instantiation of this reference architectural model to classes of autonomous vessels, with varying degree of autonomy;
- The identification of functional and operational requirements for autonomous vessels within the architectural model;
- The identification and analysis of the interdependencies and interconnections of the cyber-physical components of the C-ES.

The remainder of the paper is structured as follows: In section 2 the related work is briefly reviewed. Section 3 briefly reviews the MAF and presents the proposed extension. In Section 4 we demonstrate the use of the extended MAF to create architectural instances for variants of the C-ES, by identifying the functional and operational requirements of the C-ES; the CPSs comprising the C-ES; and the interdependencies and interconnections among them. Finally, Section 5 summarizes our conclusions and indicates directions for future research.

2 Related work

Reference architectures have been developed for the smart grid [6], [21]; service oriented architectures [26]; Industries 4.0 (RAMI4.0) [38]. In the maritime domain, the MITS [29] architecture describes the ICT components in the maritime industry and it has been used in [30] to describe the architecture of the unmanned merchant ship. A limitation of this model is the use of the OASIS

[26] reference model to identify the vessel's Operational Technology (OT) infrastructure. The IMO has proposed its e-navigation architecture, covering mostly ship to ship communications, the relationships and the sharing of information between stakeholders [37]. However, this architecture pertains exclusively to conventional ships; hence it cannot be directly applied to the autonomous ship case. ARKTRANS [24] is a reference architecture framework which captures responsibilities, relations, and dependencies in the transport sector. The European project Maritime Navigation Information Services (MarNIS) [23] has adopted the aforementioned framework. This captures the overall conceptual, logical, and technical aspects of the maritime sector. Yet, the framework is inappropriate for the C-ES case, as it is unable to capture technical operations which are crucial to understanding the operational objectives of the C-ES.

Little published work on the architecture of the C-ES exists. The MUNIN project developed a reference model of the architecture of the unmanned merchant ship [31]. The developed architecture is based on the MITS [29] architecture and on the OASIS [26] reference model. Further, [12] describes an architecture of the autonomous ship that considers only the connectivity of systems, ending up with a communication architecture. In [13] an autonomous ship architecture is proposed, based on IT components, without however taking into account the OT infrastructure.

3 The extended Maritime Architectural Framework

The MAF is a domain specific architectural methodology that was developed to overcome the challenge to coordinate the development of new systems between technology issues, governance aspects and users between existing architectures in the maritime domain. As such, the MAF establishes clear relationships between technical systems, users and related governance aspects. Similarly with other approaches, the MAF is divided into two parts, namely the multidimensional cube that provides a graphical representation of the underlying maritime domain and the examined system architecture; and a methodology to structure the examined system including the system requirements and (possible) use cases in a consistent way. The methodology is composed of three main steps leading to enable an easy mapping of system architectures to the MAF-Cube. The scope of this process is to structure the system engineering phases starting from planning over the identification of requirements to the use case development in a harmonized and formal way. This allows the user to map the results, to visualize them in the MAF-Cube, to explore interoperability issues, and to identify spots which need to be standardized [36].

The main element of the MAF is the multidimensional cube, that combines different viewpoints to provide a graphical representation of the underlying maritime domain and the examined system architecture. The cube captures three dimensions, alias *axes*, namely interoperability; hierarchical; and topological. The topological axis represents the logical location where a technology component is located. The interoperability axis addresses communication, data and

information, usage and context of a maritime system. The hierarchical axis substructures management and control systems of the maritime domain, for example for maritime transportation systems from the traffic management of a coastal area down to the radar echo of a vessel [36]. Each axis breaks down to a number of *layers*. The layers of the topological axis (ships; link; shore) are derived from IMO's breakdown of the maritime domain [25]. The layers of the interoperability axis (Regulation Governance; function; information; communication; component) cover organizational, informational and technical aspects and include the different levels of interaction (operational, functional, technical and physical) as stated in IMO's e-Navigation vision [16]. Finally, the layers of the hierarchical axis (Fields of activity; operations; systems; technical services; sensors actuators; transport objects) cover economic, technical and physical issues of a maritime system.

Information, technology, and people are crucial elements of the C-ES ecosystem [11], and the MAF is able to capture these elements. Therefore, the MAF can in principle be used for representing and analyzing the C-ES ecosystem. However, the MAF in its current form cannot capture specific characteristics of autonomous vessels; some modifications are required in order to describe the new concepts and technologies. Specifically, the topological axis should be extended to include the C-ES, the SCC, and the Link between them. This should reflect the integration of new concepts, technologies, and operational models across all the components of the interoperability and hierarchical axes. The aforementioned extensions are described below:

- **C-ES layer:** Representing the ship-side entities such as the vessel's infrastructure, operational and functional goals, processes, and systems.
- **SCC layer:** Entities of the shore side infrastructure are represented along with processes, and systems which are vital for the C-ES's operation and facilitate the interaction with in/out of maritime sector entities.
- **Link layer:** Represents the telecommunication methods and protocols between C-ES and SCC.

This conceptual extension includes important aspects of the C-ES's ecosystem. The resulting extended MAF is shown in Figure 1.

4 Putting the extended MAF in action

4.1 C-ES Functional and Operational requirements

Identifying functional and operational requirements is the first step towards modelling the C-ES ecosystem. Functional requirements support the actions of the vessel systems, whilst operational requirements support the business and organizational requirements of the C-ES ecosystem.

The operational and functional requirements of autonomous vessels have been examined in the literature. DNV-GL in [35] clarified the main functional requirements for the conventional ship. [5] has identified the functional requirements for the remote and autonomous ships, focusing on the system specifications. Further,

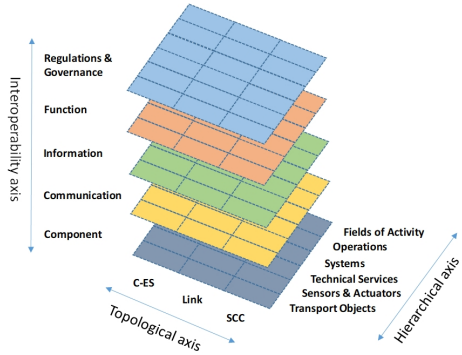


Fig. 1: Extended MAF

Table 1: Functional Requirements

Functions	Description
System functions	The necessary system functions to facilitate the ship's voyage (e.g. engine functions).
Collision avoidance	The avoidance of collision with manned objects, physical obstacles, and marine animals.
Search and Rescue	The provision of the necessary assistance to other ships or persons which are in danger at sea.
Technical Reliability	The assurance of the operations, functions, and maintenance of the C-ES's systems.
Voyage planning	The C-ES conducts route planning, determines its position, course, and speed and follows a predefined route.
Keep general lookout	The C-ES promotes its situational awareness of the area surrounding the vessel.
Cyber-Security	The C-ES's infrastructure is protected against cyber-attacks.
Physical Security	The C-ES protects its infrastructure, cargo and humans from physical attacks.

[31] analyzed the functional requirements of a merchant ship. The functional requirements of six main autonomous vessel systems have been analyzed in [34]. Additionally, the navigational, vessel engineering, and communication functions of autonomous vessels have been described in [10]. Although functions and operations described in [35] are included in [31], a set of the functions described in [5] could not be included in this classification. Considering these works and by leveraging the MAF, we identify the functional and operational requirements for the C-ES as depicted in Tables 1 and 2 respectively.

4.2 Cyber-Physical Systems of the C-ES

In order to use the extended MAF to analyze the variants of the C-ES deriving from the four autonomy levels, and in particular in order to analyze the system and component layers of the Hierarchical and Interoperability axis, we need to identify and classify the C-ES's CPSs. In [18] we identified the CPSs of the C-ES; these are shown in Figure 2. Specifically, the C-ES ecosystem comprises three

Table 2: Operational Requirements

Operations	**Description**
Navigation	Ensuring ship navigation during the voyage.
Control	the SCC is able to intervene at any time in order to control various ship's operations and functions.
Weather	The ship must be capable to operate under harsh weather conditions.
Mooring	The C-ES should be able to secure its location in permanent anchor location in the water.
Enter a port	The C-ES should be able to secure its location in the port's infrastructure.
Fail to safe	In case of emergency, the ship must stop its operations.
Rendezvous	Under specific circumstances, the crew should proceed onboard the vessel.
Transport cargo	The C-ES should have the appropriate infrastructure in order to transport cargo securely and safely.
Load/unload cargo	The C-ES should have the appropriate infrastructure in order to load/unload securely and safely cargo.
Transport people	The C-ES should have the appropriate infrastructure in order to transport passengers securely and safely.
Communication	The C-ES must establish powerful communication networks within its infrastructure and with external actors.
Passenger utilities	The C-ES must have adequate infrastructure to serve passenger's needs.
Environmental observations	The C-ES has to use contemporary sensors in order to increase its situational awareness.
Anchoring	The C-ES has to anchor in ports or anywhere under the supervision of the SCC.
Ensure seaworthiness	The C-ES must comply to the corresponding legal framework for its operations.
Maintain personnel and environmental safety	The C-ES should identify potential risks related to the safety of the crew and its environment.
Maintain preparedness	The C-ES should develop and maintain resilience-aware activities to mitigate risks and to increase its situational awareness.
Human Resources	The C-ES should ensure the management of the connected human resources systems.
Third parties relationships	The C-ES should manage relationships with suppliers, vendors, and other entities that influence its operational environment.
Cyber Security	The C-ES should follow cybersecurity standards and procedures and enforce the necessary security policies.

different classes following the MAF classification, namely the Vessel, the SCC, and the Link.

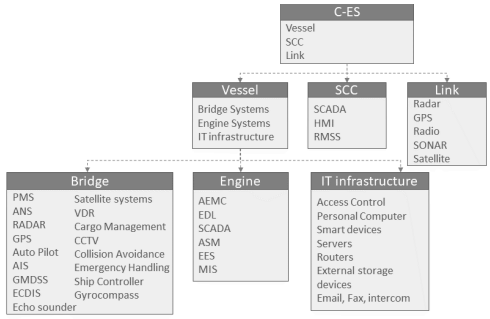


Fig. 2: C-ES's Cyber-physical systems

4.3 C-ES architectural instances

We used the extended MAF to analyze the variants of the C-ES deriving from the four autonomy levels; the result of this analysis for the hierarchical axis is presented in Table 3. Note that AL1 and AL2 are merged in this table, as both represent a remotely controlled vessel.

By examining Table 3 we conclude that AL1-AL2 and AL3 share all fields of activity among them and with the AL0, with the exception of the communication with a SCC, which is not relevant for AL0 vessels. The operations of the vessels remain the same for all ALs. The *systems* row captures the integrated systems. Although vessels belonging to AL1-AL3 inherit the systems of AL0, advanced decision support and remote control systems are introduced. These are depicted in Figure 2.

Additional technical services of the remotely controlled and of the autonomous ship respectively have been identified. CPSs identified in the previous layer reflect the technical services of each ship variant and therefore services are increasing as more CPSs are included in the infrastructure. The sensors and actuators installed in the conventional ship's infrastructure accommodate simple vessel functions, such as AutoPilot and weather observations. On the other hand, remotely controlled and autonomous ships will be equipped with advanced sensors systems able to facilitate functions such as docking, mooring, and engine maintenance. Finally, the transport objects (e.g. cargo and humans) for all vessel variants remain unaltered.

Table 4 contains the result of the analysis of the C-ES along the interoperability axis of the MAF cube. Many regulations and guidelines have been established for the AL0 ships depending on their type and fields of activity. Regulations such as [15],[22],[8] are applicable to different types of cargo ships (e.g., container, ferries, and tanker). The analysis of the regulations regarding the AL1-AL3 reveals that there is no established legal framework which marks boundaries of their operations, functions, and fields of activity[20]. Nevertheless, a lot of effort has been put on the development of guidelines from classification societies such as

Table 3: MAF Hierarchical axis for the C-ES ecosystem

C-ES: Functions			
	AL0	**AL1-AL2**	**AL3**
Fields of activity	Communication with authorities	Communication with authorities	Communication with authorities
	Ensure seaworthiness	Ensure seaworthiness	Ensure seaworthiness
	Systems to handle port operations	Systems to handle port operations	Systems to handle port operations
	Vessel Traffic service (VTS)	Vessel Traffic service (VTS)	Vessel Traffic service (VTS)
	-	Communication with SCC	Communication with SCC
Operations	Navigation	Navigation	Navigation
	Docking	Docking	Docking
	Mooring	Mooring	Mooring
Systems	Automatic Identification System (AIS)	Automatic Identification System (AIS)	Automatic Identification System (AIS)
	Electronic Chart Display and Information System (ECDIS)	Electronic Chart Display and Information System (ECDIS)	Electronic Chart Display and Information System (ECDIS)
	Global Maritime Distress and Safety System (GMDSS)	Global Maritime Distress and Safety System (GMDSS)	Global Maritime Distress and Safety System (GMDSS)
	Personnel safety systems	Personnel safety systems	-
	-	Remote maneuvering System	Remote maneuvering System
	-	Collision avoidance system	Collision avoidance system
	-	Autonomous Navigation System (ANS)	Autonomous Navigation System (ANS)
Technical services	Broadcast AIS data	Broadcast AIS data	Broadcast AIS data
	Fire protection	Fire protection	Fire protection
	Power generation	Power generation	Power generation
	Load/unload cargo	Load/unload cargo	Load/unload cargo
	-	Broadcast control commands	Broadcast control commands
	-	Sensors data fusion	Sensor data fusion
	-	-	AEMC
	-	-	Decision making
Sensors/Actuators	Auto Pilot	Auto Pilot	Auto Pilot
	Weather sensors	Weather sensors	Weather sensors
	Traffic sensors	Traffic sensors	Traffic sensors
	-	Docking actuators	Docking actuators
	-	Engine sensors/actuators	Engine sensors/actuators
Transport object	Humans	Humans	Humans
	Container	Container	Container

DNV-GL [10], Lloyd's Register[27], China classification society [7], and Beureu Veritas [34]. The functions of the sensors and actuators between different autonomy levels are differentiated, since the complexity of the sensor infrastructure of AL1-AL3 vessels is increased. The information exchange between the ship variants differs with the autonomy level. Specifically, AL1-AL3 ships rely heavily on the information of sensors and actuators since advanced systems such as collision avoidance and decision making demand high information accuracy. The *communication* plane of the MAF can capture different protocols between sensors and actuators. AL0 ships usually employ protocols such as Modbus and radio signals, while AL1-AL3 ship communications may be established by leveraging contemporary communication protocols such as ZigBee, WiFi and Satellite connections. The *components* plane exhibits diversity in different autonomy levels. In particular, the autopilot, weather sensors and other environmental analysis sensors are crucial for AL0 ships, whilst contemporary engine actuators, navigation and docking sensors are vital for AL1-AL3 vessels.

4.4 Interconnections, Dependencies and Interdependencies among CPS

To complete the architectural description of the C-ES, the interconnections, dependencies and interdependencies among the CPSs need to be also identified. Two CPSs are interconnected when there exists information exchange between

Table 4: MAF Interoperability axis for the C-ES ecosystem

C-ES: Sensors & Actuators			
	AL0	**AL1-AL2**	**AL3**
Regulations	COLREGs	Could be adopted from conventional	-
	NMEA 2000		-
	Directive 2010/65/EU		-
Functions	Navigation	Navigation	Navigation
	Environment monitoring	Environment monitoring	Environment monitoring
	Temperature, speed and vibration measurements	Temperature, speed and vibration measurements	Temperature, speed and vibration measurements
	-	Mooring	Mooring
	-	Berthing	Berthing
Information	State/value of collision avoidance sensors	State/value of collision avoidance sensors	State/value of collision avoidance sensors
	State/value of steering sensors	State/value of steering sensors	State/value of steering sensors
	State/value of engine room sensors	State/value of engine room sensors	State/value of engine room sensors
	Distance from the port	Distance from the port	Distance from the port
	Depth of sea	Depth of sea	Depth of sea
	-	Objects at sea	Objects at sea
Communication	Modbus	Modbus	Modbus
	Satellite Com	Satellite Com	Satellite Com
	Radio (VHF)	Radio (VHF)	Radio (VHF)
	-	WiFi	WiFi
Components	Auto Pilot	Auto Pilot	Auto Pilot
	Weather sensors	Weather sensors	Weather sensors
	Temperature, speed and vibration sensors	Temperature, speed and vibration sensors	Temperature, speed and vibration sensors
	-	-	Docking actuators
	-	Engine actuators	Engine actuators
	-	Depth sounders	Depth sounders
	-	-	Navigation sensors and actuators

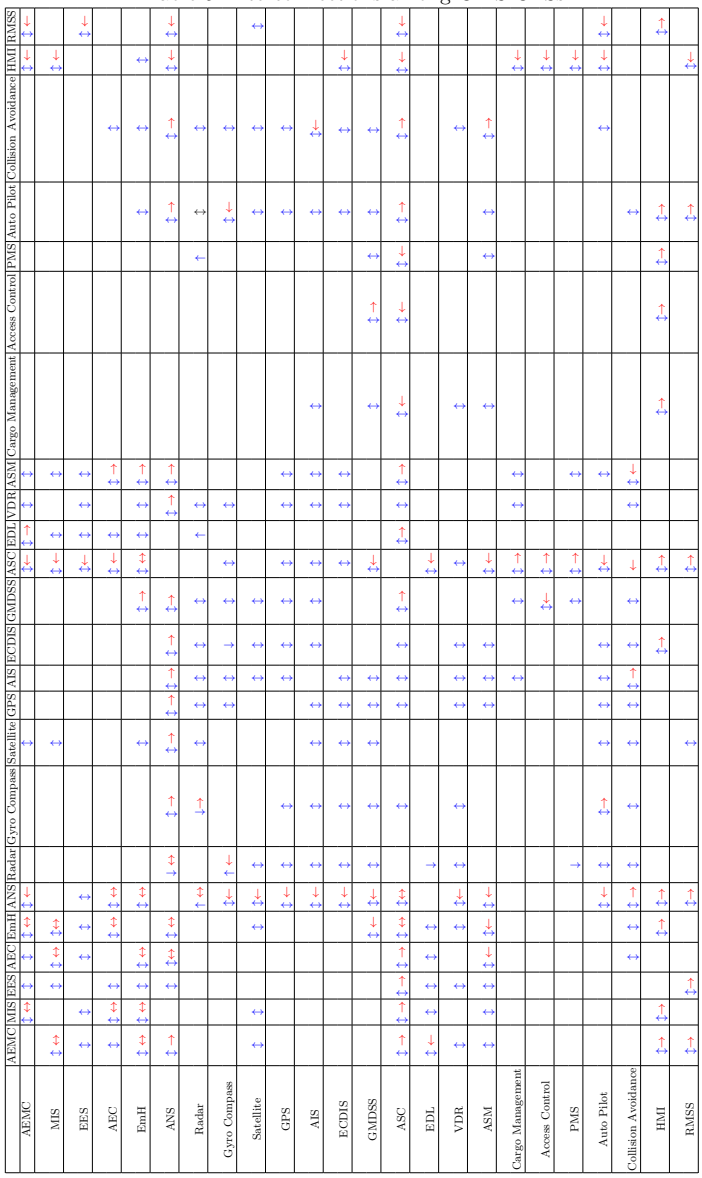
them; when two CPSs are connected and the state of one system influences the state of the other, the systems are dependent. Two systems are interdependent when there exists bilateral dependency between them.

Table 5 depicts the interconnections along with the control flows within the CPSs of the C-ES. In particular, the data and control flows for each system are represented with blue arrows and red arrows respectively.

The C-ES's CPSs are all complex components in which changes may occur as a result of operational and/or functional processes. This complexity derives from the combination of IT and OT systems, the size of the C-ES ecosystem, the diversity of the installed components, and the different fields of activity. According to [28] an effective way to examine complex systems is to view them as a group of interacting systems. Accordingly, we examine the dependencies and interdependencies of the C-ES's systems considering three main groups of systems; the Bridge Automation System (BAS), the Engine Automation System (EAS), and the SCC. Furthermore, the dependencies and interdependencies of the three critical onboard components [18], namely the AIS; the ECDIS; and the GMDSS, all subsystems of the BAS are depicted in Figures 4,5, and 6.

Figure 3 represents the systems that can directly or indirectly be affected by potential systems state's changes. Figure 4 depicts the dependencies and interdependencies of the AIS, Figure 5 those of the ECDIS, and Figure 6 those of the GMDSS.With an eye towards identifying the most critical CPSs, and understanding the impact propagation among them, by way of considering their interconnections, dependencies and interdependencies, we first map the information in Table 5 onto two graphs, whose nodes represent CPSs and edges represent connections. We then employ certain graph analysis metrics, that were calcu-

Table 5: Interconnections among C-ES CPSs



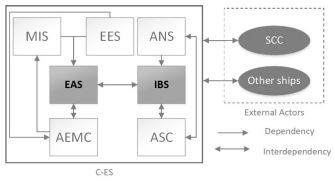


Fig. 3: General Ecosystem

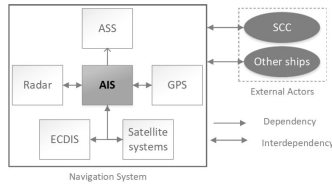


Fig. 4: AIS

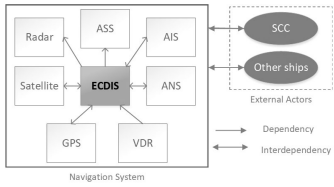


Fig. 5: ECDIS

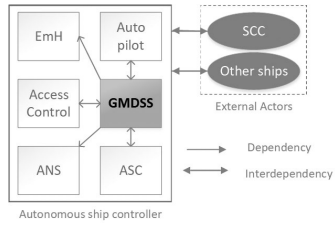


Fig. 6: GMDSS

lated by leveraging the CASOS ORA tool from Carnegie Mellon University [1], to analyze the systems' criticality.

The exponential ranking centrality (ERC) defines the centrality of each system as its trustworthiness; it is based on the degree of trust that other systems have in it; the AIS and the ECDIS have the highest ERC value (1 and 0,985 respectively). The Betweenness Centrality (BC) metric allows the identification of the systems which hold the most critical position considering the connections and interconnections. The higher the value of the BC of a system, the more systems are connected to it and therefore a potential failure would affect the whole system.Finally, The degree centrality estimates the number of connections a system has. In particular, the Total Degree (TD) is the sum of the links in and from the systems. A system with high TD is a well connected node; therefore its operations and functionalities can influence other systems and, in case of failure, may provoke bigger damage to the infrastructure.

The aforementioned analysis identified the most critical CPSs of the C-ES, taking into account the trustworthiness (ERC), the percentage of the paths that pass through each system (BC), and the number of the connections that each CPS has (TD). The analysis focused on both connections/interconnections and dependencies/interdependencies of the systems. The ANS and the Autonomous Ship Controller (ASC) have the highest values as it can be seen in Table 6. This denotes that a potential failure of such systems could provoke a sequence of failures among CPSs and therefore increase the impact to the ship. Additionally, according to Table 6, the AIS and the ECDIS are the most trustworthy. Thus, a malfunction of these can lead to cascading effects on other ship's systems.

Table 6: Graph analysis results

CPS	ERC	CPS	BC	CPS	Total Degree
AIS	1	**ASC**	0.161	**ASC**	0.891
ECDIS	0.985	ANS	0.051	ANS	0.739
GPS	0.933	GMDSS	0.049	ASM	0.609

5 Conclusions

A central trend within the digital transformation of the maritime industry is increased autonomy of vessels. This needs to be supported by engineering specifications, regulations, standards, etc. This, in turn, necessitates the existence of an architectural framework that will facilitate the specification, implementation, and operation of such vessels. In this paper we extended the MAF to allow the representation of autonomous vessels; we used this reference architecture to define instances of cyber-enabled ships; we mapped functional and operational requirements of such systems on the reference architecture; and we identified and anlyzed the dependencies and interconnections of cyber-physical systems that comprise a cyber-enable ship. We intend to use this reference architecture and the results obtained herein, along with an appropriate requirements engineering method, to elicit cybersecurity requirements for the cyber-enabled ship.

References

1. CASOS. http://www.casos.cs.cmu.edu/index.php, accessed: 2019-09-10
2. e-Navigation. http://www.imo.org/en/OurWork/Safety/Navigation/Pages/eNavigation.aspx, accessed: 2019-10-10
3. Smart grid reference architecture. Technical report, CEN-CENELEC-ETSI Smart Grid Coordination Group (2012)
4. A technical specification for the common shore-based system architecture (cssa). Technical report, International Association of Marine Aids to Navigation and Lighthouse Authorities (2015)
5. Bergström, M., Hirdaris, S., Valdez B., O., Kujala, P., Sormunen, O., Lappalainen, A.: Towards the unmanned ship code (06 2018)
6. CEN-CENELEC-ETSI Smart Grid Coordination Group: Smart grid reference architecture. Tech. rep. (2012)
7. China classification society: Guidelines for autonomous cargo ships 2018. Tech. rep. (2018)
8. Council of European Union: Regulation (EC) No 725/2004 (2004)
9. Cross, J., Meadow, G.: Autonomous ships 101. Journal of Ocean Technology **12**, 23–27 (2017)
10. DNVGL: Autonomous and remotely operated ships, class guideline. Tech. rep. (2018)
11. Fitton, O., Prince, D., Germond, B., Lacy, M.: The future of maritime cyber security. Tech. rep. (2015)
12. Höyhtyä, M., Huusko, J., Kiviranta, M., Solberg, K., Rokka, J.: Connectivity for autonomous ships: Architecture, use cases, and research challenges. In: Information and Communication Technology Convergence (ICTC), 2017 Int. Conference on. pp. 345–350. IEEE (2017)

13. Im, I., Shin, D., Jeong, J.: Components for smart autonomous ship architecture based on intelligent information technology. Procedia computer science **134**, 91–98 (2018)
14. International Maritime Organization : IMO takes first steps to address autonomous ships. http://www.imo.org/en/mediacentre/pressbriefings/pages/08-msc-99-mass-scoping.aspx (2018), [Online; accessed 24-May-2019]
15. International Maritime Organization: Convention on the international regulations for preventing collisions at sea (1972)
16. International Maritime Organization: Msc 85/26/add.1, annex 20 strategy for the development and implementation of e-navigation. Tech. rep. (2009)
17. Systems and software engineering — Architecture description. Standard, International Organization for Standardization (2011)
18. Kavallieratos, G., Katsikas, S., Gkioulos, V.: Cyber-attacks against the autonomous ship. In: 4th Workshop on the Security of Industrial Control Systems of Cyber-Physical Systems (CyberICPS 2018), Barcelona, Spain, pp. 20-36, 2019, Computer Security. pp. 20–36. Springer International Publishing, Cham (2019)
19. Kitada, M., et al.: Command of Vessels in the Era of Digitalization. In: Int. Conference on Applied Human Factors and Ergonomics. pp. 339–350. Springer (2018)
20. Komianos, A.: The autonomous shipping era. operational, regulatory, and quality challenges. TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation **12**, 335–348 (01 2018). https://doi.org/10.12716/1001.12.02.15
21. National Institute of Standards and Technology : Introduction to nistir 7628 guidelines for smart grid cyber security. Tech. rep. (2010)
22. National Marine Electronics Association (NMEA): Nmea 2000 standard (1972)
23. Natvig, M.: Final report on the marnis e-maritime architecture. Tech. rep. (2008)
24. Natvig, M., Westerheim, H., Christiansen, I.: Arktrans the norwegian system framework architecture for multimodal transport systems supporting freight and passenger transport (06 2019)
25. NCSR 1-28: Report to the maritime safety committee, international maritime organization, sub-committee on navigation communications and search and rescue. Tech. rep. (2014)
26. OASIS: Reference architecture foundation for service oriented architecture. Tech. rep. (2009)
27. Register, Lloyds: Cyber-enabled ships: Deploying information and communications technology in shipping–lloyds register's approach to assurance. London: Lloyds Register, http://www. marinelog. com/index. php (2016)
28. Rinaldi, S.M., Peerenboom, J.P., Kelly, T.K.: Identifying, understanding, and analyzing critical infrastructure interdependencies. IEEE Control Systems Magazine **21**(6), 11–25 (Dec 2001). https://doi.org/10.1109/37.969131
29. Rødseth, Ø.J.: e-maritime standardisation requirements and strategies (2009), http://www.mits-forum.org/architecture.html
30. Rødseth, Ø.J., Tjora, Å.: A system architecture for an unmanned ship. In: Proceedings of the 13th Int. Conference on Computer and IT Applications in the Maritime Industries (COMPIT) (2014)
31. Rødseth, O.J., Tjora, A., Baltzersen, P.: Munin d4.5: Architecture specification. Tech. rep. (2013)
32. Tran, T., M., N.: Integrating requirements of industry 4.0 into maritime education and training: case study of vietnam (2018)
33. Uslar, M., et al.: Applying the smart grid architecture model for designing and validating system-of-systems in the power and energy domain: a european perspective

34. Veritas, Bureau: Guidelines for autonomous shipping. Tech. rep. (2017)
35. Vindøy, V.: A functionally oriented vessel data model used as basis for classification. In: 7th Int. Conference on Computer and IT Applications in the Maritime Industries, COMPIT. vol. 8 (2008)
36. Weinert, B., Hahn, A., Norkus, O.: A domain-specific architecture framework for the maritime domain. Informatik 2016 (2016)
37. Weintrit, A.: Development of the imo e-navigation concept–common maritime data structure. In: Int. Conference on Transport Systems Telematics. pp. 151–163. Springer (2011)
38. ZVEI Die Elektroindustrie: Reference architecture model industrie 4.0. Tech. rep. (2015)

6 Article II: Managing Cyber Security Risks of the Cyber-Enabled Ship [2]

Journal of
Marine Science and Engineering

MDPI

Article

Managing Cyber Security Risks of the Cyber-Enabled Ship

Georgios Kavallieratos *, Sokratis Katsikas *

Department of Information Security and Communication Technology, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

* Correspondence: georgios.kavallieratos@ntnu.no (G.K.); sokratis.katsikas@ntnu.no (S.K.)

Received: 12 September 2020; Accepted: 28 September 2020; Published: 30 September 2020

Abstract: One aspect of the digital transformation process in the shipping industry, a process often referred to as Shipping 4.0, is the increased digitization of on board systems that goes along with increased automation in and autonomy of the vessel. This is happening by integrating Information Technology with Operation Technology systems that results in Cyber Physical Systems on which the safe operations and sailing of contemporary and future vessels depend. Unavoidably, such highly interconnected and interdependent systems increase the exposure of the vessel's digital infrastructure to cyber attacks and cyber security risks. In this paper, we leverage the STRIDE and DREAD methodologies to qualitatively and quantitatively assess the cyber risk of Cyber Physical Systems on board digitalized contemporary and future ships. Further, we propose appropriate cyber security baseline controls to mitigate such risks, by applying a systematic approach using a set of criteria that take into account the security requirements; the cyber risks; the possible attacks; and the possibly already existing controls, to select from the list of controls provided in the Industrial Control Systems (ICS) overlay of the NIST Guide to ICS Security. The results are expected to support the decision-making and the design of a security architecture for the cyber-enabled ship.

Keywords: cyber-enabled ship; cyber risk assessment; cyber security controls selection; cyber physical systems

1. Introduction

Despite the fact that today almost all ships are to some extent digitalized, the shipping industry addresses the digital transformation challenge, including the emergence of crew-less vessels [1]. Such vessels come in two broad categories, namely the remotely operated vessel and the autonomous vessel; both kinds are referred to as *cyber-enabled ships (C-ES)* [2]. The C-ES is a cyber physical ecosystem which consists of the vessel itself, a Shore Control Center (SCC) that controls and handles the C-ES, the communication links between the vessel and the SCC, and other ships in the vicinity.

The integration of Information Technology (IT) and Operation Technology (OT) to form *Cyber Physical Systems (CPS)*, which constitute a central element of the digital transformation process in many application domains is unavoidably accompanied by an increase and a diversification of the cyber risks that the domain is facing. This is mainly due to the fact that whereas traditional operations were designed with no need for cyber security in mind, modern IT-enabled operations are allowed to be accessed and controlled by outward-facing information systems, through interfaces that are rarely adequately secure [3].

The C-ES is no exception. Although most of the C-ES CPSs are parts of today's conventional ships, their exposure to contemporary technologies, aiming to be controlled and monitored remotely, increases the attack surface and makes them more vulnerable to cyber-attacks. Indeed, research on the cyber security risks of autonomous and unmanned vessels [2,4] has revealed an increased

attack surface and several vulnerable systems. Thus, ship-side cyber security incidents, such as, for example, the ones reported in Reference [5–7] , have already occurred; in fact, such incidents have been increasing at an alarming rate over the last three years [8]. Such incidents may also impact the safety of humans, operations, and cargo.

In the light of these findings, of the increased financial value of the sector [9], and of the multitude of potential attackers, including such with advanced capabilities, the promotion of cyber security and safety of the C-ES ecosystem becomes very important [10]. The first step towards strengthening the cyber security posture of an ecosystem is to understand, analyze, and manage the cyber risks that it faces; this will eventually drive the design of a security architecture that includes appropriate cyber security controls that will mitigate the risks.

Risk is defined as "the effect of uncertainty on objectives" [11]. Cyber Security risk is associated with the potential that threats will exploit vulnerabilities of an asset or group of assets and thereby cause harm to an organization. Cyber risk is assessed in terms of the likelihood of a *threat*[1] occurring, the extent of the *vulnerabilities*[2] to the threat, and the magnitude of the *impact*[3]; these constitute the *elements* of cyber risk.

The risk management process as specified in ISO 31000[13] comprises five sub-processes [11], as shown in Figure 1:

1. The external and internal context for cyber security risk management should be established, which involves setting the basic criteria necessary for cyber security risk management, defining the scope and boundaries, and establishing an appropriate organization operating the cyber security risk management.
2. Risks should be assessed, i.e., identified, quantified or qualitatively described, and prioritized against risk evaluation criteria and objectives relevant to the organization.
3. Controls to reduce, retain, avoid, or share the risks should be selected and a risk treatment plan defined.
4. Information about risk should be exchanged and/or shared between the decision-makers and other stakeholders.
5. Risks and their elements should be monitored and reviewed to identify any changes at an early stage and to maintain an overview of the complete risk picture. This is why, as Figure 1 illustrates, the cyber security risk management process can be iterative for risk assessment and/or risk treatment activities.

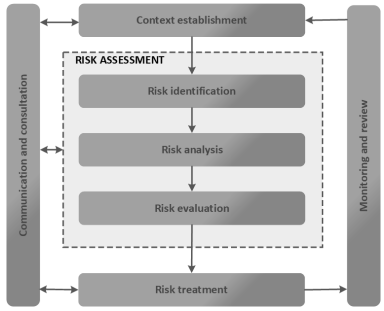


Figure 1. Risk management process.

[1] A threat is the potential cause of an unwanted incident, which may result in harm to a system or organization [12].
[2] A vulnerability is a weakness of an asset or control that can be exploited by one or more threats [12].
[3] Impact or consequence is the outcome of an event affecting objectives [12]

In this paper, we focus on the risk assessment and the risk treatment sub-processes. Risk assessment methods are quantitative, qualitative, or semi-quantitative. Quantitative risk assessment is based on using mathematical methods and rules and assigns a numerical value, often in the [1-x] range to each risk. The results are less subjective than those of the other two types, and therefore drive the process of control selection more effectively, but they cannot be easily communicated to non-technically oriented decision-makers. In contrast, qualitative risk assessment is based on applying non-numerical methods and assigns a level value to each risk, such as low, medium, and high. This type of assessment has a limited number of results, but these are more comprehensible to decision-makers. Finally, semi-quantitative risk assessment combines rules and methods for evaluating the risk by combining numeric values and levels; for example, the [1-x] range can easily be converted into qualitative expressions that help risk communication to decision-makers.

STRIDE and DREAD have been selected for the work described herein. These methods can effectively analyze highly interconnected CPSs comprising heterogeneous components [14], and they are most appropriate for analyzing systems under development. In such systems, the operational and functional requirements are not established yet. Alternative approaches need such requirements to produce valid results. In contrast, STRIDE and DREAD facilitate the analysis of conceptual systems by answering questions regarding the security objectives of the targeted ecosystem. Moreover, the combination of qualitative and quantitative methods to analyze the cyber risk provides a holistic view, not captured by other methods. Further, this hybrid approach facilitates the communication of the results to relevant stakeholders while allowing the representation of cyber risk in numeric form, thus facilitating the assessment of the effectiveness of controls at later stages of the risk treatment process. Finally, both STRIDE and DREAD are being widely used in both academia and industry [15].

Risk treatment is the process followed to modify risk [11]. A risk can be treated by :

- *modifying* its level, by introducing controls;
- *retaining* it, with no further action taken;
- *avoiding* it, by avoiding the activity or condition that gives rise to the particular risk;
- *sharing* it with other party or parties, for example, by means of insurance and/or risk financing.

The four options for risk treatment are not mutually exclusive. Sometimes a combination of options, such as modifying risks and sharing or retaining any residual risks, can be beneficial.

Individual elements of the cyber risk of, as well as *attacks* [4] against individual CPSs in the C-ES, have been studied, and proposals for risk assessment approaches have appeared in the literature. However, to the best of our knowledge, a holistic assessment of the cyber risks of the whole CPS part of the C-ES ecosystem, comprising all of the aforementioned types of risk assessment methods, which leads to concrete proposals for cyber security controls and can also be used by non-technical decision-makers, has not been made available.

In this paper:

- we extend our previous work in Reference [2] on qualitative risk assessment of CPSs on board the C-ES to all CPSs identified in Reference [16];
- we provide a quantitative risk assessment for all C-ES CPSs identified in Reference [16];
- we propose an approach for systematically selecting appropriate cyber security controls to mitigate the cyber risks; and
- we demonstrate the workings of the approach by applying it to select cyber security controls for the most vulnerable CPSs on board the C-ES.

The remainder of the paper is structured as follows: In Section 2, we review the relevant literature. In Section 3, we use the STRIDE method [17] as modified in Reference [2] to analyze the threats and the

---

[4] An attack is an attempt to destroy, expose, alter, disable, steal, or gain unauthorized access to or make unauthorized use of an asset [12]. An attack is a particular way of a threat to exploit one or more vulnerabilities.

attack scenarios for the CPSs of the C-ES that have been identified in Reference [16] and to qualitatively assess the related risks. In Section 4, we turn our attention to quantitatively assessing the risks, by leveraging a variant of the DREAD method [18] adapted for use in CPSs. Our proposed approach for systematically selecting cyber security controls is presented in Section 5, where also its workings are demonstrated by means of applying it to select controls for the three most vulnerable on-board CPSs of the C-ES. Finally, Section 6 summarizes our conclusions and indicates directions for future research.

2. Related Work

A wealth of cyber risk assessment methods applicable to general purpose IT systems exists. Whilst these can be and have been applied to IT systems in the maritime domain, they cannot accurately assess cyber risks related to CPSs [19]. Cyber risk assessment methods for CPSs more often than not are domain specific, as they need to take into account safety as an impact factor additional to the "traditional" impact factors of confidentiality, integrity, and availability [3]. In the maritime domain, a review of cyber security risk assessment methods appeared in Reference [20]. Rødseth et al. in Reference [21] proposed a risk assessment method for the unmanned merchant ship. Although the method aims to identify both safety and security risks, particular focus is given on hazard identification and to the accordant risks, with cyber security left largely unaddressed. Tam et al. in Reference [4] proposed the MaCRA model-based framework for maritime cyber-risk assessment and applied it to a number of example scenarios [22]. However, the aim of MaCRA is not to assess the risks or flaws of specific systems, but rather to facilitate the understanding of cyber risks in the maritime domain. B. Svilicic et al. in Reference [23] proposed a framework for assessing cyber risks in ships and applied it to the case of the Electronic Chart Display and Information System (ECDIS).

Several works in the literature have analyzed security threats and risks for specific systems used in specific types of autonomous and remotely controlled vessels. Among these, Bolbot et al. in Reference [24] identified and analyzed safety related cyber-attacks in an autonomous inland ferry; their analysis covers safety aspects regarding the navigational and propulsion system of the ferry. Silverajan et al. in Reference [25] explored security issues and cyber attacks targeting systems of smart ships. Awan et al. in Reference [26] have analyzed 59 documented accidents to better understand the vulnerabilities of Integrated Bridge System (IBS) components. Svilicic et al. in Reference [27] present a study on the cyber security resilience of a shipboard Integrated Navigational System (INS) installed on a RoPax ship engaged in international trade. Wang et al. in Reference [28] propose a secure relative integrated navigation method to counteract injected fault measurement attacks. Balduzzi et al. in Reference [29] presented a security evaluation of the Automatic Identification System (AIS), by introducing threats affecting both the implementation in online providers and the protocol specification. Lund et al. in Reference [30] described a proof-of-concept attack on an INS and its integrated ECDIS, and demonstrated the attack on a vessel. Kavallieratos et al. in Reference [2] identified potential cyber attack scenarios and qualitatively evaluated the accordant risks for a number of CPSs of the C-ES ecosystem, both on-board and in the SCC.

Systematic methods for selecting security controls for IT systems either view the problem of control selection as an investment problem and apply management tools and financial analysis to optimize the selection [31], or in the context of responding to an intrusion, i.e., when a specific attack has been already detected as taking place [32]. A combinatorial optimization model to efficiently select security controls was proposed in Reference [31]. However, security control selection is still largely performed empirically, particularly for CPSs. In the maritime domain, potential cyber security controls for systems on board autonomous and remote controlled vessels have also been proposed. Bothur et al. in Reference [33] discussed the security vulnerabilities that smart ships face, and described security countermeasures, particularly procedural and technical solutions, by following a defense in depth approach. Silverajan et al. in Reference [25] analyzed the main systems of an unmanned smart ship and proposed defense strategies against previously discussed cyber attacks and threats. Bolbot et al. in Reference [24] analyzed safety-related cyber attacks for the navigational and propulsion

systems, evaluated the accordant risks and proposed general security recommendations. Sahay et al. in Reference [34] proposed an SDN framework to mitigate cyber attacks and improve the resilience in the smart ship's communication network. None of the above works followed a systematic, risk-based process for selecting the controls. Further, the aforementioned analyses focused on defense strategies and controls that are not system-specific.

3. Qualitative Risk Assessment

3.1. STRIDE

STRIDE is an acronym formed by the initials of six security threats: *S*poofing, *T*ampering, *R*epudiation, *I*nformation disclosure, *D*enial of Service, and *E*levation of privileges. Spoofing is the capability of an adversary to pretend that they are someone or something else. Tampering is the alteration or disruption of aasset of the system, e.g., disk, network, or memory. Repudiation is someone's allegation that they did not do something which influences the system's operation or were not responsible for the results of their actions. Information disclosure reveals confidential information to unauthorized entities. Denial of Service reduces the availability of the system by, e.g., exhausting system resources. Elevation of Privilege is an adversary's ability to assume privileges that allow them to execute unauthorized actions.

The method was developed by Loren Kohnfelder and Praerit Garg in 1999 and is described in detail by A. Shostack in Reference [17]. Security threats are analyzed and attack scenarios are developed in light of the security objectives of *Authenticity, Integrity, Non-repudiation, Confidentiality, Availability, and Authorization*. STRIDE can be used to discover potential threats and vulnerabilities as early as the design phase. Therefore, it enables the analysis of systems that are under development, thus facilitating the requirements engineering elimination process and adherence to security-by-design principles [35]. STRIDE has been used in ecosystem environments similar to the C-ES, where CPSs are prominent [14,36,37].

3.2. STRIDE for the CPSs of the C-ES Ecosystem

STRIDE is a threat modeling method. In our previous work [2] we proposed a modified version of STRIDE and used it to model threats, to develop cyber attack scenarios, and to qualitatively assess the accordant risks for fourteen CPSs of the C-ES ecosystem, namely the Engine Automation System (EAS), the Bridge Automation System (BAS), the Shore Control Center (SCC), the Autonomous Engine Monitoring and Control System (AEMC), the Engine Efficiency System (EES), the Maintenance Interaction System (MIS), the Navigation Systems (NavS), the Autonomous Ship Controller (ASC), the Human-Machine Interface (HMI), the Remote Maneuvering Support System (RMSS), the Emergency Handling system (EmH), the Automatic Identification System (AIS), the Electronic Chart Display and Information System (ECDIS), and the Global Maritime Distress and Safety System (GMDSS). A reference architecture for the C-ES was proposed in Reference [16], in which five CPSs additional to those in the architecture proposed in Reference [2] were identified, namely the Collision Avoidance (C.A.), Radar, CCTV, Advanced Sensor Module (ASM), and Auto Pilot (AP) systems.

The results of the application of the modified STRIDE of Reference [2] to these systems, as well as to the Voyage Data Recorder (VDR), Cargo Management, and Engine Data Logger (EDL) systems that, due to space limitations, were not reported in Reference [2] are presented in Tables A1–A8 in the Appendix. In these tables "I" stands for "Impact", "L" stands for "Likelihood" and "R" stands for "Risk". Three distinct values have been assigned to the impact and the risk: Low (L), Medium (M), and High (H). The possible values for the likelihood of a cyber attack are: Very Likely (VL), Moderate (M), and Rare (R). These values have been assigned by applying the criteria that are described in Tables 1 and 2, and in Figure 2 of Reference [2], and are summarized in Table 1. The values have been determined by both consulting the literature and by leveraging the authors' own expertise.

Table 1. Impact and likelihood criteria.

	Impact Criteria
High	Significant financial damage to the shipping company; or physical damage to the infrastructure; or loss of human life.
Medium	Financial damage to the shipping company; or disruption of operations; or legal sanctions; or breach of the confidentiality, integrity or availability of information.
Low	Delay of non-critical operations; or breach of the confidentiality, integrity or availability of non-sensitive information.
	Likelihood Criteria
Very Likely	Existence of highly motivated and capable attackers and no controls in place; or wide availability of exploits; or high exposure of the system to the internet.
Moderate	Existence of highly motivated and capable attackers and inadequate controls in place; or wide availability of exploits that require physical access; indirect exposure of the system to the internet.
Rare	Absence of highly motivated and capable attackers; or adequate controls in place; no exposure of the system to the internet.

4. Quantitative Risk Assessment

4.1. DREAD

DREAD [18] stands for *D*amage, *R*eproducibility, *E*xploitability, *A*ffected users/systems, and *D*iscoverability. *Damage* represents the damage that a cyber attack may inflict to the system; along with the *Affected Users/Systems*, it represents the *Impact* of the attack. *Reproducibility* represents the ability of the attacker to reproduce the attack, whilst *Exploitability* their ability to exploit the system's vulnerabilities and to carry out the attack. *Discoverability* represents the capacity of the adversary to identify system's vulnerabilities. The sum of Reproducibility, Exploitability, and Discoverability represents the *Likelihood* of the cyber attack.

STRIDE and DREAD are interrelated and provide a systematic analysis of novel systems to ensure the security of such systems early in the design phase. The former facilitates the qualitative security analysis of the system by considering six security threats that violate the corresponding security objectives. The latter quantifies the identified risks that result by the attack scenarios developed with STRIDE.

4.2. DREAD for the CPSs of the C-ES Ecosystem

Quantitative risk analysis aims to assign meaningful numbers to elements of risk analysis; impact and likelihood are such elements. Assessing the cyber risk by considering the probability of an attack occurring results in rating numbers and values that can cause confusion and disagreement among stakeholders in the risk management process [18]. DREAD aims to overcome such limitations by quantifying specific aspects (Damage potential, Reproducibility, Exploitability, Affected systems, and Discoverability) of security threats and attacks to assign meaningful numbers to the elements of risk by means of Formulas (1) and (2).

Building upon the analysis of the security threats and the corresponding attack scenarios for the CPSs of the C-ES as reported in Reference [2] and in Section 3.2 above, DREAD is used to produce quantitative estimates of the risks of the identified attack scenarios. The risk value is calculated by using the following formulas:

$$Impact = \frac{\sum(Damage, Affectedsystems)}{2}, \quad (1)$$

$$Likelihood = \frac{\sum(Reproducibility, Exploitability, Discoverability)}{3}, \quad (2)$$

$$Risk = \frac{(Impact + Likelihood)}{2}. \quad (3)$$

The values for the DREAD components are determined according to the criteria shown in Table 2, which have been adapted from Reference [18] so as to include CPSs aspects. These criteria are analyzed in Reference [38].

Table 2. DREAD (*D*amage, *R*eproducibility, *E*xploitability, *A*ffected users/systems, and *D*iscoverability) criteria [38].

	High (3)	**Medium (2)**	**Low (1)**
D	The adversary is able to bypass security mechanisms; get administrator access; upload/modify the CPS content.	Leakage of confidential information of the CPSs (functions/source code); cause partial malfunction/disruption of the system.	Leaking non-sensitive information; the attack is not possible to extend to the other CPSs on-board.
R	The cyber-attack can be reproduced anytime to the targeted CPS.	The adversary is able to reproduce the attack but under specific risk conditions.	Although the attacker knows the CPS's vulnerabilities/faults, s/he is unable to perform the cyber-attack.
E	The cyber-attack can be performed by a novice adversary in a short time.	A skilled adversary may launch the attack.	The attack requires an extremely skilled person and in-depth knowledge of the targeted CPS.
A	All CPSs are affected	Partial users/systems, non-default configuration	The attack affects only the targeted CPS.
D	The CPS's vulnerabilities are well known and the attacker is able to get access to the relevant information to exploit them.	The CPS's vulnerabilities/faults are not well known and the adversary needs to get access to the CPS.	The threat has been identified and the vulnerabilities have been patched.

Tables 3 and 4 depict the resulting risk value of each CPS for each STRIDE threat, calculated according to the Formulas (1)–(3), and by both consulting the literature, and by leveraging the authors' own expertise.

Table 3. Cyber risks in engine and Shore Control Center (SCC) Cyber Physical Systems (CPSs).

	EAS	**AEMC**	**EDL**	**ASM**	**EES**	**MIS**	**SCC**	**RMSS**	**HMI**
S	1.33	1.75	1.5	2.25	2	1.5	2.05	1.75	2.16
T	1.67	1.5	1.25	1.28	1.75	2.25	1.67	1.5	2.16
R	1.25	1.25	1.25	1.25	1	1.25	1.42	1.25	1.25
I	1.42	1	1.25	1.66	1.25	1.5	1.42	1.75	2
D	2	1.5	1.25	2	1.75	1.75	2.05	1.75	2.16
E	1.26	1.25	1.25	1.25	1.5	1.5	1.25	1.5	1.5

Table 4. Cyber-risks in bridge CPSs.

	BAS	**AIS**	**ECDIS**	**GMDSS**	**ASC**	**ANS**	**EmH**	**C.A.**	**Radar**	**VDR**	**Cargo**	**CCTV**	**AP**
S	1.83	2.33	2.42	2.25	2.17	1.92	1.25	1.91	2.25	1.5	1.5	2.16	1.5
T	1.67	2.42	2.17	2.5	2.5	1.92	1.25	2.08	2.08	1.5	1.5	1.83	1.75
R	1.25	2.33	1.25	1.5	1.75	1.5	1	1.25	1.66	1.25	1.5	1.5	1.25
I	1.83	2.33	2.33	2.25	1.75	1.75	1.25	1.41	1	1.5	1.75	1.91	1.5
D	2	2	2.5	2.5	2.58	1.92	1.5	1.91	2	1.5	1.25	1.91	1.75
E	1.25	1.92	2.33	2.17	2	2.17	1	1.25	1.5	1.5	1.5	1.75	1.5

4.3. Discussion

As already mentioned in the introduction, a semi-quantitative risk assessment facilitates the communication of risks to non-technical decision-makers. In this case, expressing the results of the quantitative risk assessment in Section 4.2 will also allow comparisons to be made between these and those of the qualitative risk assessment in Section 3.2. To this end, the risk values in Tables 3 and 4 can be converted to qualitative risk levels as follows:

Low: DREAD risk ≤ 1
Medium: $1 <$ DREAD risk ≤ 2
High: $2 <$ DREAD risk ≤ 3

Table 3 suggests that Spoofing and Denial of Service are the most critical threats both among the engine room and the SCC systems. Similarly, Table 4 suggests that the Spoofing, Tampering, and Denial of Service threats present the highest risk levels among the bridge systems of the C-ES. Tampering and Information disclosure are medium risk threats, and Repudiation and Elevation of privileges are low risk threats.

Moreover, a single risk value for each examined system can be assigned, equal to the largest among the risk values for the same system. Table 5 depicts these numerical values, as well as the results of the quantitative risk assessment converted to qualitative according to the rules above and those of the qualitative risk assessment.

Table 5. Quantitative versus qualitative risks.

CPS	DREAD	Quantitative Risk Analysis	Qualitative Risk Analysis
ECDIS	2.5	High	High
GMDSS	2.5	High	High
ASC	2.5	High	High
AIS	2.42	High	High
MIS	2.25	High	Medium
ASM	2.25	High	Medium
Radar	2.25	High	Medium
ANS	2.17	High	High
HMI	2.16	High	High
CCTV	2.16	High	Medium
C.A.	2.08	High	Medium
SCC	2.05	High	Medium
EES	2	Medium	Medium
BAS	2	Medium	Medium
EAS	2	Medium	Medium
RMSS	1.75	Medium	Medium
AEMC	1.75	Medium	Medium
CaMa	1.75	Medium	Medium
EmH	1.5	Medium	Medium
VDR	1.5	Medium	Medium
EDL	1.5	Medium	Medium
AP	1.75	Medium	Medium

It can be noticed that none of the studied CPSs faces low risk, and that the risk levels determined by the qualitative and the quantitative risk assessment methods for most of these systems are similar; deviations should be attributed to the increased subjectivity of the qualitative risk assessment. Despite the deviations, both approaches suggest that the navigational systems are among the most vulnerable on-board CPSs of the C-ES.

In previous work [16], we analyzed the interconnections and interdependencies among the CPSs of the C-ES. By leveraging these results along with the quantitative risks depicted in Tables 3 and 4, the propagation of risks among the CPSs can be examined. Note, for example, which the AIS is

interconnected and interdependent with the ECDIS, the Radar, and the ASM, systems that also face the highest risk values. This is because systems which are interconnected and interdependent share similar security risks, because they inherit the vulnerabilities of the most vulnerable CPSs which can be used as intermediate stepping stones for launching attacks [38].

5. Cyber Risk Treatment

The ISO27005 risk management approach aims at identifying risk treatment strategies rather than designing the security architecture of the system under study. A necessary prerequisite for designing such an architecture for the C-ES is to select appropriate controls for each individual component, and to consolidate these into a coherent and consistent whole that will take into account not only the risks, but also the requirements stemming from the C-ES's environment. Accordingly, we propose an approach for managing the risks of the C-ES, as depicted in Figure 2, where six sub-processes are specified, along with their inputs and outputs. The Environmental Analysis sub-process for the C-ES has been carried out in Reference [16]; the Threat Analysis sub-process has been carried out in Reference [2]; and the Security Requirements Elicitation sub-process has been carried out in Reference [39]. In this work we focus on the Cyber Risk Assessment sub-process (Sections 3 and 4) and on the Control Selection sub-process (Section 5.1). The Security Architecture Design sub-process is the subject of future work.

5.1. Control Selection

This activity includes the initial selection of a set of minimum security controls to protect the system based on a set of criteria that take into account the security requirements; the cyber risks; the possible attacks; and the possibly already existing controls. This set will ensure baseline protection of the system; the baseline controls are the starting point for the design of the overall security architecture, which will derive from the application of tailoring to the set of security control baselines to account for peculiarities of the system and of the organization that owns or operates the system. In the sequel our approach for selecting the set of baseline controls is described.

A number of sources (e.g., Reference [40–42]) provide sets of security controls from which a selection can be made. All of these sources pertain to information systems rather than cyber-physical systems; hence their applicability in the case under study is limited. However, Appendix G of the NISTGuide to Industrial Control Systems (ICS) Security [43] provides the *ICS overlay*, which is a partial tailoring of the controls and control baselines in Reference [41,42], which adds supplementary guidance specific to ICS. We will be using this source to select controls from, according to the following set of criteria, adapted from Reference [44]:

C1: Kind of CPS that needs to be protected;
C2: Security aspects that need to be protected.
C3: Threats that need to be eliminated.
C4: Potential control alternatives.
C5: The value of the CPS to protect, according to its importance. This has been assessed within the process of attack path analysis, performed in Reference [38].
C6: The likelihood of threat occurrence. This derives from the threat analysis performed within the risk assessment process of Sections 3 and 4.
C7: Risk coverage provided by alternative controls.

As an example, the values of the control selection criteria for the spoofing threat against AIS are as follows:

C1: Navigational CPS;
C2: Integrity and availability. These are derived from the security requirements that have been established in Reference [39].
C3: Spoofing/Tampering/DoS. These derive from the threat analysis results performed in Reference [2] and in Sections 3 and 4.
C4: Encryption/Tamperproof hardware.
C5: High. This has been assessed within the process of attack path analysis, performed in Reference [38].
C6: Very likely. This derives from the threat analysis performed within the risk assessment process of Sections 3 and 4.
C7: Low. No alternative controls are already in place.

and lead to selecting the IA-3 control category of Reference [43]. An example of a control that belongs to this category is the establishment and use of an authentication infrastructure for such devices, such as, e.g., the one proposed in Reference [45,46].

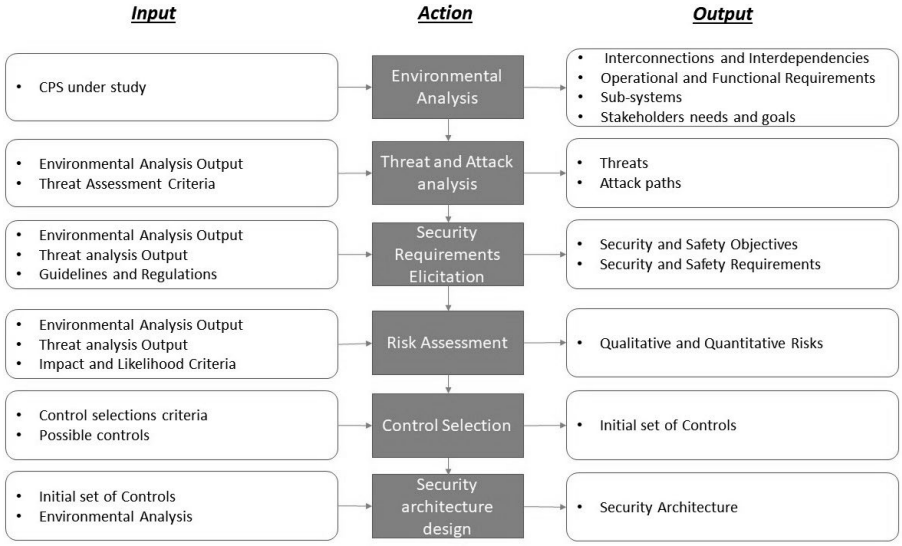


Figure 2. Overall control selection approach.

5.2. Application to the Case of the AIS, the ECDIS, and the GMDSS

The results of the application of the process described above to the three most vulnerable on-board systems of the C-ES are shown in Tables 6–8.

Table 6. Control selection for the Automatic Identification System (AIS).

Threat	Risk	Requirement	Objective	Control Category
Spoofing	Medium	Reliable authentication mechanisms must be in place in order to uniquely identify the actors that read, modify, or transmit AIS data, as well as to authenticate the system itself and its services.	Authentication	Device Identification and Authentication (IA-3)
Tampering	High	The confidentiality and integrity of the data exchanged between internal (on board) systems and external actors (SCC or other vessel) should be ensured by appropriate mechanisms depending on the actors and the type of the data in transit.	Confidentiality/ Integrity	Port and I/O Device Access (SC-41), Software Firmware and Information Integrity (SI-7)
Repudiation	High	The AIS should implement the security services in order to protect the system from loss of control or possession of information.	Possession and Control, Non-repudiation	Device Identification and Authentication (IA-3), Physical Access Control (PE-3), Monitoring Physical Access (PE-6 (1)), Account Management (AC-2 (2),(3)), Non-repudiation (AU-10), Information System Component Inventory (CM-8 (4))
Information Disclosure	High	Voyage data, such as destination port or cargo related information, should be confidential to prevent potential leakage to adversaries.	Confidentiality, Integrity	Cryptographic Key Establishment and Management (SC-12 (1)), Cryptographic Protection (SC-13)
Denial of Service	Medium	The connectivity between system and external actors and between on board systems must be continuous.	Availability, Utility	Internal System Connections (CA-9), Information System Backup (CP-9 (1), (2), (3), (5)), Power Equipment and Cabling (PE-9), Denial of Service Protection (SC-5)
Elevation of Privileges	High	The AIS must be able to implement lock mechanisms (e.g., lock HMI screen) upon request by the administrator or after a configurable time of idleness.	Authenticity, Non- repudiation	Internal System Connections (CA-9), Monitoring Physical Access (PE-6)

Table 7. Control selection for the Electronic Chart Display and Information System (ECDIS).

Threat	Risk	Requirement	Objective	Control Category
Spoofing	High	The use of ECDIS must be restricted only to authorized and well trained personnel.	Authenticity, Integrity	Device Identification and Authentication (IA-3), Port and I/O Device Access (SC-41), Time Stamps (AU-8), Plan of Action and Milestones (CA-5)
Tampering	Medium	The ECDIS must be able to control the flows of voyage-related data sent to other ships and to the SCC.	Integrity, Authenticity	Device Identification and Authentication (IA-3), Audit Review Analysis and Reporting (AU-6 (3), (6)), Plan of Action and Milestones (CA-5)
Repudiation	Medium	The ECDIS should be able to audit sent and received data to external actors.	Integrity, Non repudiation	Internal System Connections (CA-9), Time Stamps (AU-8), Physical Access Control (PE-3), Monitoring Physical Access (PE-6 (1))
Information Disclosure	High	The confidentiality and integrity of the data exchanged between internal (on board systems and external actors (SCC or other vessel) should be ensured by appropriate mechanisms depending on the actors and the type of the data in transit.	Confidentiality	Cryptographic Protection (SC-13), Port and I/O Device Access (SC-41), Device Identification and Authentication (IA-3), Protection of Information at Rest (SC-28)
Denial of Service	High	The communication between the ECDIS and the satellite system should be continuously available.	Availability	Internal System Connections (CA-9), Incident Handling (IR-4 (4)), Denial of Service Protection (SC-5)
Elevation of Privileges	Low	The use of ECDIS must be restricted only to authorized and well trained personnel	Possession and Control	Device Identification and Authentication (IA-3), Unsuccessful Logon Attempts (AC-7)

Table 8. Control selection for the the Global Maritime Distress and Safety System (GMDSS).

Threat	Risk	Requirement	Objective	Control Category
Spoofing	High	Distress signals transmitted through the GMDSS must be verified by external actors, such as SCC and other ship's subsystems, such as the Autonomous Engine Monitoring and Control (AEMC) and Navigation systems	Confidentiality, Authenticity	Continuous Monitoring (CA-7 (1)), Time Stamps (AU-8)
Tampering	High	The signals transmitted to external actors or subsystems must be appropriately encrypted	Integrity	Cryptographic Protection (SC-13)

Table 8. *Cont.*

Threat	Risk	Requirement	Objective	Control Category
Repudiation	Medium	The authenticity of the transmitted GMDSS signals and data in transit to the Autonomous Ship Controller (ASC), to other subsystems, and to the SCC must be ensured	Authenticity, Non repudiation	Physical Access Control (PE-3), Access Control for Output Devices (PE-5), Unsuccessful Log on Attempts (AC-7)
Information Disclosure	High	The measures to protect the confidentiality and integrity of data should not downgrade their utility	Confidentiality	Continuous Monitoring (CA-7(1))
Denial of Service	Medium	Safety signals transmitted through the GMDSS to other on board systems and external actors must be continuously available.	Availability	Internal System Connections (CA-9), Incident Handling (IR-4 (4)), Contingency Plan (CP-2), Denial of Service Protection (SC-5)
Elevation of privileges	Medium	The ASC must be able to provide security, safety, and dynamic data to the GMDSS, when needed	Authenticity, Possession and Control	Device Identification and Authentication (IA-3)

Table 9 depicts the consolidated controls per studied CPS.

Table 9. Baseline controls.

CPS	Baseline Controls
AIS	Device Identification and Authentication (IA-3), Port and I/O Device Access (SC-41), Software Firmware and Information Integrity (SI-7 (1)), Cryptographic Protection (SC-13), Tamper Protection (PE-3(5)), Physical Access Control (PE-3), Monitoring Physical Access (PE-6 (1)), Account Management (AC-2 (2),(3)), Non-repudiation (AU-10), Information System Component Inventory (CM-8 (4)), Cryptographic Key Establishment and Management (SC-12 (1)), Internal System Connections (CA-9), Information System Backup (CP-9 (1), (2), (3), (5)), Power Equipment and Cabling (PE-9), Denial of Service Protection (SC-5)
ECDIS	Device Identification and Authentication (IA-3), Port and I/O Device Access (SC-41), Time Stamps (AU-8), Plan of Action and Milestones (CA-5), Audit Review Analysis and Reporting (AU-6 (3), (6)), Internal System Connections (CA-9), Physical Access Control (PE-3), Monitoring Physical Access (PE-6 (1)), Cryptographic Protection (SC-13), Protection of Information at Rest (SC-28), Incident Handling (IR-4 (4)), Denial of Service Protection (SC-5), Unsuccessful Logon Attempts (AC-7)
GMDSS	Continuous Monitoring (CA-7 (1)), Time Stamps (AU-8), Cryptographic Protection (SC-13), Physical Access Control (PE-3), Access Control for Output Devices (PE-5), Unsuccessful Log on Attempts (AC-7) , Internal System Connections (CA-9), Incident Handling (IR-4 (4)), Contingency Plan (CP-2), Denial of Service Protection (SC-5), Device Identification and Authentication (IA-3).

Some of these controls are recommended for all systems (Device Identification and Authentication (IA-3), Cryptographic Protection (SC-13), Denial of Service Protection (SC-5), Physical Access Control (PE-3), Internal System Connections (CA-9)), whilst others are recommended for two or for only one of the studied systems. During the security architecture design phase, the controls identified for all systems will need to be re-considered, consolidated, checked for applicability in the specific environment, conformance to guidelines, compliance to standards etc.

As is typical with risk treatment strategies, the application of security controls does modify (reduce) the risk but does not eradicate it. To complete the risk treatment process one needs to assess the effectiveness of the applied controls, to consider the residual risk within the specific environmental and organizational context and to possibly repeat the process until the residual risk falls below the

accepted risk level. This process can be effectively performed when the whole security architecture of the C-ES has been determined; accordingly, this is an item for future work.

One of the distinctive characteristics of CPSs is their ability to interconnect dynamically, sometimes to address scope beyond the originally intended one. This often results in emergent, hence unpredictable, behavior. In order to effectively secure CPSs in such situations, dynamic assessment of cyber risk is recommended. The proposed methodology, as it now stands, cannot capture such behavior. However, it can be extended, along the lines followed in Reference [36].

6. Conclusions

We systematically analyzed the cyber security risks of the CPSs of the C-ES. Both a qualitative and a quantitative assessment of these risks was undertaken, by using the STRIDE and DREAD methods respectively. By leveraging the results of both assessments and applying a systematic structured approach, we identified appropriate baseline cyber security controls for each of the three more vulnerable on-board CPSs. As future work, we intend to build on these results to design the security architecture of instances of the C-ES.

Author Contributions: Conceptualization, G.K. and S.K.; methodology, G.K.; investigation, G.K.; writing—original draft preparation, G.K.; writing—review and editing, S.K. and G.K.; supervision, S.K.; project administration, S.K.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	linear dichroism

Appendix A. STRIDE Tables

Table A1. Collision Avoidance—C.A.

T	Collision Avoidance—C.A.	I	L	R
S	An adversary may spoof the existence of another ship in the vicinity, thereby causing the vessel to change its route.	H	M	H
T	Tampering of sensor data may cause the vessel to collide with other ships/human made obstacles/environmental obstacles.	H	M	H
R	The repudiation of actions of the collision avoidance system is unacceptable since all such actions are clearly defined and assigned to the necessary equipment.	M	R	L
I	The leakage of information exchanged via the collision avoidance system may reveal information regarding the position of the vessel and its voyage.	L	M	M
D	A disruption of the operation of the collision avoidance system may cause physical damage.	H	M	H
E	An adversary with high privileges may disrupt the normal operation of the system.	H	R	M

Table A2. Radar.

T	Radar	I	L	R
S	An attacker may spoof the identity of a ship in the vicinity and confuse the ANS to deviate from the intended route.	M	VL	H
T	An attacker may violate the integrity of the dynamic data of the Radar (positioning data) and cause physical damage to the vessel.	M	M	M
R	The dynamic data sent by the Radar can be spoofed, rendering other systems unable to identify the source of the data.	M	R	L
I	No confidential or sensitive data are transmitted through Radar.	L	R	L
D	A signal jamming of the Radar may cause disruption on the services and confusion to the other systems regarding the position and the speed of the vessel.	M	VL	H
E	An attacker with high administrative access is able to turn off the Radar or alter the transmitted data.	M	R	L

Table A3. CCTV.

T	CCTV	I	L	R
S	An adversary may spoff the identity of a monitoring camera in the engine room and confuse the EAS's decision-making.	M	M	M
T	An attacker is able to alter the images depicted in the monitoring system and cause damage. The integrity of the data sent from CCTV is crucial since they contribute to the situational awareness of the C-ES.	M	M	M
R	Wrong data regarding the vessel's environment could be sent to the ANS. The ANS cannot perform any integrity check or identify the malicious source of the system's data.	M	R	L
I	Potential leakage of the data exchanged between CCTV and SCC may cause GDPR violations and hence financial and legal damages to the shipping company.	H	M	H
D	Any disruption of the system's services may lead to loss of the vessel's situational awareness.	H	M	H
E	An adversary with administrative access to the system may disable the cameras on-board or the access control systems and hence violate the authenticity and availability of information within the vessel.	H	R	M

Table A4. VDR.

T	VDR	I	L	R
S	An adversary may pretend the identity of the legitimate ECDIS and store wrong/malicious data to the VDR.	H	R	M
T	By leveraging the weak encryption of the stored data, the attacker may change voyage/dynamic/static data and confuse the decision-makers.	H	M	H
R	The data are stored automatically to the VDR by a well-defined process; such a threat is not applicable.	M	R	L
I	An attacker may gain access to unauthorized data by leveraging the weak encryption of the data and the absence of an access control mechanism.	H	M	H
D	The adversary may disrupt the storage of data service by sending data for storage continuously.	M	M	M
E	Potential access to the systems as an administrator could cause damage to the stored data, such as delete, alter, or leak confidential data.	H	R	M

Table A5. Cargo management.

T	Cargo Management	I	L	R
S	An adversary may spoof the identity of the cargo or ship owner and gain access to sensitive data regarding the type of the cargo or the destination port.	H	R	M
T	Tampering of the data derived from the cargo monitoring system may cause damage to the cargo.	H	M	H
R	An attacker may confuse the cargo management process by attacking the CCTV system and sending malicious data, such as fire on the deck, pirates on-board, etc.	H	R	M
I	The violation of the data confidentiality of the cargo management system may lead to GDPR violation and cause financial damage.	H	M	H
D	An adversary may disrupt the operation of the system by attacking the communication line between ship and shore, thus making the cargo handling service unavailable.	H	M	H
E	An attacker with administrative access to the cargo management system may cause damage to the cargo (cargo loss), financial damage to the shipping company, or damage to the reputation of the shipping company.	H	R	M

Table A6. Engine Data Logger—EDL.

T	EDL	I	L	R
S	An adversary may assume the identity of the captain/system administrator by logging in with the credentials of the administrator and gain access to the engine related data.	H	M	H
T	The attacker may alter the data stored in the EDL, such as the engine performance data and cause damage to the engine or confusion during the investigation of an accident.	H	M	H
R	An adversary may log wrong data to the EDL by leveraging the lack of control actions to properly track the logged-in users.	H	R	M
I	The information and data stored in EDL are not confidential; therefore a potential leakage cannot cause significant damage to the vessel.	L	M	L
D	The disruption of the system's operation may cause physical damage to the engine room by confusing the MIS to proceed with the actions foreseen in case of engine failure.	M	R	M
E	An attacker with high administrative rights may change the system's configuration and cause violations of data integrity and/or availability.	H	R	M

Table A7. Advanced Sensor Module—ASM.

T	Advanced Sensor Module—ASM	I	L	R
S	An adversary may gain access to the ASM by deploying a malware. By leveraging the malware, the attacker is authenticated as system administrator and therefore fault messages could be sent to other on board systems, such as navigation and engine monitoring systems. This scenario may cause damage to the ship and/or financial damage to the company.	H	M	H
T	An attacker may tamper the engine sensor data transmitted to ANS and provide fake measurements (e.g., temperature, engine oil).	H	R	M
R	The repudiation of the actions of the ASM is unacceptable since its functions are based on automated and well-defined process. Potential violation of the repudiation of the system may cause confusion in the decision-making process.	H	R	M
I	Potential data leakage of the ASM or potential disclosure of the sensor architecture facilitates the reconnaissance stage of a cyber-attack.	M	M	M
D	An adversary may flood the systems with fake data, thus affecting its ability to share the valid data with the engine and navigational systems. The disruption of the system's operation may cause significant damage to the vessel and/or financial damage to the shipping company since the vessel's situational awareness capability will be adversely affected.	H	M	H
E	Due to the weak access control in the ASM, an adversary may gain system access with high administrative rights and disrupt the ASM operation and/or services.	H	R	M

Table A8. Auto Pilot—AP.

T	Auto Pilot—AP	I	L	R
S	An attacker may spoof the identity of the Shore Control Center and provide wrong position coordinates to the AP.	M	M	M
T	The alteration of the AP data may lead to the grounding of the vessel and cause financial and physical damage.	H	R	M
R	If the source of the received data is spoofed, the AP will not be able to identify the system that sent the wrong data.	H	R	M
I	An adversary may gain access to information related to the vessel's position or the destination port. Financial damage may result due to the leakage of confidential information.	M	R	L
D	The attacker may send fake data continuously to the AP and hence disable its normal operation.	H	M	H
E	An attacker with administrative rights is able to change the vessel's route and cause financial damage.	H	R	M

References

1. Cross, J.; Meadow, G. Autonomous ships 101. *J. Ocean Technol.* **2017**, *12*, 23–27.
2. Kavallieratos, G.; Katsikas, S.; Gkioulos, V. Cyber-attacks against the autonomous ship. In *Proceedings of the SECPRE 2018, CyberICPS 2018. Lecture Notes in Computer Science, Vol 11387*; Springer Nature Switzerland: 2018; pp. 20–36.
3. BIMCO and CLIA and ICS and INTERCARGO and INTERMANAGER and INTERTANKO and IUMI and OCIMF and WORLD SHIPPING COUNCIL. In *The Guidelines on Cyber Security Onboard Ships*; Technical Report; 2018.
4. Tam, K.; Jones, K. Cyber-risk assessment for autonomous ships. In Proceedings of the International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Oxford, UK, 3–4 June 2018; pp. 1–8.
5. USCG. Cyber Incident Exposes Potential Vulnerabilities Onboard Commercial Vessels. Available online: https://www.dco.uscg.mil/Portals/9/DCO%20Documents/5p/CG-5PC/INV/Alerts/0619.pdf (accessed on 02.09.2020).
6. Jones, M. Spoofing in the Black Sea: What Really Happened? Available online: https://www.gpsworld.com/spoofing-in-the-black-sea-what-really-happened/ (accessed on 02.09.2020).
7. MARAD. 2019-012-Persian Gulf, Strait of Hormuz, Gulf of Oman, Arabian Sea, Red Sea-Threats to Commercial Vessels by Iran and Its Proxies. Available online: https://www.maritime.dot.gov/content/2019-012-persian-gulf-strait-hormuz-gulf-oman-arabian-sea-red-sea-threats-commercial-vessels (accessed on 02.09.2020).
8. Cyber Attacks on Maritime OT Systems Increased 900% in Last Three Years. 2020. Available online: https://safety4sea.com/cyber-attacks-on-maritime-ot-systems-increased-900-in-last-three-years/#:~:text=Cyber%2Dattacks%20on%20the%20maritime,security%20firm%20Naval%20Dome%20reveals (accessed on 29.08.2020).
9. Kessler, G.; Craiger, J.; Haass, J. A Taxonomy Framework for Maritime Cyber Security: A Demonstration Using the Automatic Identification System. *Transnav Int. J. Mar. Navig. Saf. Sea Transp.* **2018**, *12*, 429.
10. Katsikas, S.K. Cyber security of the autonomous ship. In Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security, Abu Dhabi, UAE, 2 April 2017; pp. 55–56.
11. International Organization for Standardization, ISO. *ISO/IEC 27005:2018 Information Technology—Security Techniques—Information Security Risk Management*; ISO Geneva, Switzerland: 2018.
12. International Organization for Standardization, ISO. *ISO/IEC 27000:2018(en) Information Technology—Security Techniques—Information Security Management Systems—Overview And Vocabulary*; ISO Geneva, Switzerland: 2018.
13. International Organization for Standardization, ISO. *ISO 31000:2018 Risk management—Guidelines*; ISO Geneva, Switzerland: 2018.
14. Kavallieratos, G.; Gkioulos, V.; Katsikas, S.K. Threat analysis in dynamic environments: The case of the smart home. In Proceedings of the 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), Santorini Island, Greece, 29–31 May 2019; pp. 234–240.

15. Hussain, S.; Kamal, A.; Ahmad, S.; Rasool, G.; Iqbal, S. Threat modelling methodologies: A survey. *Sci. Int.* **2014**, *26*, 1607–1609.
16. Kavallieratos, G.; Katsikas, S.; Gkioulos, V. Modelling Shipping 4.0: A Reference Architecture for the Cyber-Enabled Ship. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Phuket, Thailand, 23–26 March 2020; pp. 202–217.
17. Shostack, A. *Threat Modeling: Designing for Security*; John Wiley & Sons: 2014.
18. Microsoft. Chapter 3—Threat Modeling. 2010. Available online: https://docs.microsoft.com/en-us/previous-versions/msp-n-p/ff648644(v=pandp.10)?redirectedfrom=MSDN (accessed on 25.08.2020).
19. Ali, S.; Al Balushi, T.; Nadir, Z.; Hussain, O.K. Risk Management for CPS Security. In *Proceedings of Cyber Security for Cyber Physical Systems*; Springer International Publishing: 2018; pp. 11–34.
20. You, B.; Zhang, Y.; Cheng, L.C. Review on Cyber Security Risk Assessment and Evaluation and Their Approaches on Maritime Transportation. In Proceedings of the 30th Annual Conference of International Chinese Transportation Professionals Association, 2017.
21. Rødseth, Ø.J.; Burmeister, H.C. Risk assessment for an unmanned merchant ship. *Transnav Int. J. Mar. Navig. Saf. Sea Transp.* **2015**, *9*, 357–364, doi:10.12716/1001.09.03.08.
22. Tam, K.; Jones, K. MaCRA: A model-based framework for maritime cyber-risk assessment. *WMU J. Marit. Aff.* **2019**, *18*, 129–163.
23. Svilicic, B.; Kamahara, J.; Celic, J.; Bolmsten, J. Assessing ship cyber risks: A framework and case study of ECDIS security. *WMU J. Marit. Aff.* **2019**, *18*, 509–520.
24. Bolbot, V.; Theotokatos, G.; Boulougouris, E.; Vassalos, D. Safety related cyber-attacks identification and assessment for autonomous inland ships. In Proceedings of the International Seminar on Safety and Security of Autonomous Vessels (ISSAV), Aalto University, Finland, 17–20 September 2019.
25. Silverajan, B.; Ocak, M.; Nagel, B. Cyber Security Attacks and Defences for Unmanned Smart Ships. In Proceedings of the 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Halifax, NS, Canada, 30 July–3 August 2018; pp. 15–20.
26. Awan, M.; Al Ghamdi, M. Understanding the Vulnerabilities in Digital Components of an Integrated Bridge System (IBS). *J. Mar. Sci. Eng.* **2019**, *7*, 350.
27. Svilicic, B.; Rudan, I.; Jugović, A.; Zec, D. A Study on Cyber Security Threats in a Shipboard Integrated Navigational System. *J. Mar. Sci. Eng.* **2019**, *7*, 364.
28. Wang, Y.; Wang, Y.; Feng, X. Ship Security Relative Integrated Navigation with Injected Fault Measurement Attack and Unknown Statistical Property Noises. *J. Mar. Sci. Eng.* **2020**, *8*, 305.
29. Balduzzi, M.; Pasta, A.; Wilhoit, K. A Security Evaluation of AIS Automated Identification System. In Proceedings of the 30th Annual Computer Security Applications Conference, ACSAC'14, Association for Computing Machinery, New York, NY, USA, December 2014; pp. 436–445, doi:10.1145/2664243.2664257.
30. Lund, M.; Hareide, O.; Jøsok, Ø. An Attack on an Integrated Navigation System. *J. Ocean Technol.* **2017**, *12*, 23–27.
31. Schilling, A.; Werners, B. Optimal selection of IT security safeguards from an existing knowledge base. *Eur. J. Oper. Res.* **2016**, *248*, 318–327.
32. Nespoli, P.; Papamartzivanos, D.; Gómez Mármol, F.; Kambourakis, G. Optimal Countermeasures Selection Against Cyber Attacks: A Comprehensive Survey on Reaction Frameworks. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 1361–1396.
33. Bothur, D.; Zheng, G.; Valli, C. A critical analysis of security vulnerabilities and countermeasures in a smart ship system. In Proceedings of the Australian Information Security Management Conference, Perth, Australia, 5–6 December 2017.
34. Sahay, R.; Sepulveda, D.; Meng, W.; Jensen, C.D.; Barfod, M.B. CyberShip: An SDN-based Autonomic Attack Mitigation Framework for Ship Systems. In Proceedings of the International Conference on Science of Cyber Security, Beijing, China, 14–16 August 2018; pp. 191–198.
35. Sandra Domenique Zinsmaier, H.L.; Waldvogel, M. A Practical Approach to Stakeholder-driven Determination of Security Requirements based on the GDPR and Common Criteria. In Proceedings of the 6th International Conference on Information Systems Security and Privacy (ICISSP 2020), Valletta, Malta, 26 November 2020; pp. 473–480.

36. Kavallieratos, G.; Chowdhury, N.; Katsikas, S.; Gkioulos, V.; Wolthusen, S. Threat Analysis for Smart Homes. *Future Internet* **2019**, *11*, 207.
37. Seifert, D.; Reza, H. A security analysis of cyber-physical systems architecture for healthcare. *Computers* **2016**, *5*, 27.
38. Kavallieratos, G.; Katsikas, S. Attack Path Analysis for Cyber-Physical Systems. In Proceedings of the CyberICPS 2020, Guildford, UK, 12 July 2020.
39. Kavallieratos, G.; Diamantopoulou, V.; Katsikas, S. Shipping 4.0: Security requirements for the Cyber-Enabled Ship. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6617–6625.
40. Federal Office for Information Security. *IT-Grundschutz-Catalogues*; 13th Version; 2013.
41. JOINT TASK FORCE. Security and Privacy Controls for Federal Information Systems and Organizations. *NIST Spec. Publ.* **2020**, *800*, 8–13.
42. JOINT TASK FORCE. Control Baselines for Information Systems and Organizations. *NIST Spec. Publ.* **2020**, doi:10.6028/NIST.SP.800-53B-draft.
43. Stouffer, K.; Pillitteri, V.; Marshall, A.; Hahn, A. Guide to industrial control systems (ICS) security. *NIST Spec. Publ.* **2015**, *800*, 247.
44. Government of Spain, Ministry of Finance and Public Administration. *MAGERIT—Version 3.0 Methodology for Information Systems Risk Analysis and Management*; 2014; pp. 1–109.
45. Goudossis, A.; Katsikas, S.K. Towards a secure automatic identification system (AIS). *J. Mar. Sci. Technol.* **2019**, *24*, 410–423.
46. Goudosis, A.; Katsikas, S. Secure AIS with Identity-Based Authentication and Encryption. *Transnav Int. J. Mar. Navig. Saf. Sea Transp.* **2020**, *14*, 287–298, doi:10.12716/1001.14.02.03.

7 Article III: Shipping 4.0: Security requirements for the Cyber-Enabled Ship [3]

Shipping 4.0: Security requirements for the Cyber-Enabled Ship

IEEE Transactions on Industrial Informatics (Volume: 16, Issue: 10, Oct. 2020)

Georgios Kavallieratos[1], Vasiliki Diamantopoulou[2], Sokratis K. Katsikas[1,3]
[1]Dept. of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik, Norway
[2]Dept. of Information and Communication Systems Engineering, School of Engineering, University of the Aegean, Samos, Greece
[3]Faculty of Pure and Applied Sciences, Open University of Cyprus, Nicosia, Cyprus

The Cyber-Enabled Ship (C-ES) is either an autonomous or a remotely controlled vessel which relies on interconnected cyber physical-systems (CPS) for its operations. Such systems are not well protected against cyber attacks. Considering the criticality of the functions that such systems provide, it is important to address their security challenges, thereby ensuring the ship's safe voyage. In this work we leverage the Maritime Architectural Framework reference architecture to analyze and describe the environment of the C-ES. We then apply the Secure Tropos methodology to systematically elicit the security requirements of the three most vulnerable CPSs onboard a C-ES, namely the Automatic Identification System (AIS), the Electronic Chart Display Information System (ECDIS) and the Global Maritime Distress and Safety System (GMDSS). The outcome is a set of cyber security requirements for the C-ES ecosystem in general and these systems in particular.

***Index Terms*—cyber-physical systems, cyber-security, autonomous ships security, security requirements engineering**

I. Introduction

INDUSTRY 4.0 was initially coined to describe the trend towards automation and data exchange in manufacturing technologies and processes; nowadays it encompasses areas which are not normally classified as an industry, such as smart cities, for instance, and describes the trend towards increasing automation and connectivity, by leveraging technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), and Big Data Analytics regardless of domain of application, leading to the appearance of terms such as *cities 4.0*. Accordingly, the term *Shipping 4.0* was coined in 2016 to describe the new developments in digitalization of shipping, to reflect the very similar developments in land based industry which commonly goes under the name of Industry 4.0 [1].

In the maritime sector, despite the fact that nowadays almost all ships are automated in some way, the shipping industry is coming to alignment with Industry 4.0 with the emergence of crewless vessels [2]. Such vessels come in two broad categories, namely the remotely operated vessel and the autonomous vessel; both kinds are referred to as cyber-enabled ships (C-ES). The C-ES is a cyber-physical ecosystem which consists of the vessel itself, a Shore Control Center (SCC) that controls and handles the C-ES, the communication links between the vessel and the SCC, and other ships in the vicinity. The C-ES ecosystem consists of both Information Technology (IT) and Operational Technology (OT) systems which are crucial for the vessel's secure and safe operations.

The integration of Information Technology (IT) and Operation Technology (OT) that constitutes a central element of the digital transformation process in any application domain is unavoidably accompanied by an enlargement and diversification of the cyber risks that the domain is facing, with existing risks being increased and new risks being introduced. This is mainly due to the fact that whereas traditional operations were designed with no need for cyber security in mind, modern IT-enabled operations are allowed to be accessed and controlled by information systems connected to the internet, through interfaces that are rarely adequately secure.

The shipping industry and the cyber-enabled ship in particular is no exception. Although most of the C-ES CPS are parts of today's conventional ships, their exposure to contemporary technologies, aiming to be controlled and monitored remotely, increases the attack surface and makes them more vulnerable to cyber-attacks. Indeed, research on the cyber security risks of autonomous and unmanned vessels [3], [4] has revealed an increased attack surface and vulnerable systems. This enlarged attack surface has already made ship-side cyber security incidents such as, for example, the ones reported in [5], [6], [7] possible.

In the light of these findings, the increased financial value of the sector [8], and the multitude of potential attackers, including such with advanced capabilities, the promotion of cybersecurity and safety of the C-ES ecosystem is very important [9]. In order to strengthen the cyber-security posture of the ecosystem, it is necessary to define a security architecture. Acknowledging the fact that the C-ES ecosystem is characterized by high complexity and by the complex interconnections, dependencies and interdependencies among its constituent CPSs, it follows that a systematic approach needs to be followed when attempting to establish cyber security requirements, both of the ecosystem as a whole and of each individual CPS in the ecosystem.

In this paper, we first propose a security requirements elici-

Corresponding author: G. Kavallieratos (email: georgios.kavallieratos@ntnu.no).

tation process for the C-ES ecosystem. An architectural framework needs to be combined with a security requirements elicitation method to derive such requirements. The SecureTropos methodology [10] and the Maritime Architectural Framework (MAF) reference architecture [11] were identified as important elements for implementing the process. According to a threat analysis of on board systems of the C-ES [3]; a risk assessment of such systems [4]; and the known vulnerabilities of such systems [12], the Automatic Identification System (AIS), the Electronic Chart Display Information System (ECDIS), and the Global Maritime Distress and Safety System (GMDSS) have been identified as the most vulnerable on board systems of the C-ES. We then proceed with applying the process to the case of the C-ES ecosystem, and in particular to these systems. The outcome is a set of cyber security requirements for these systems, checked for their validity against the criteria specified in [13].

The remainder of the paper is structured as follows: Section II discusses related work. Section III describes our proposed security requirements elicitation process. Section IV presents the results of the application of the process to the C-ES case, and specifically the cyber-security requirements of the three most vulnerable CPSs among the C-ES systems. Finally, in Section V we summarize our conclusions and suggest areas for future work.

II. Related work

The security requirements of the autonomous vessels have only been scarcely and non-systematically examined. The technical and non-technical communication requirements for an autonomous merchant ship have been analyzed in [14]. However, the security requirements for such communications systems were not considered. The data requirements for wireless transmission of autonomous ships have been identified in [15]. Bureau Veritas in [16] described the functional requirements of six main systems of the autonomous ship, but without considering the corresponding security requirements in detail. The security requirements of a vessel's control system components have been described in [17]. Although [17] provides a comprehensive analysis of the cyber-security requirements as they derive from relevant standards, only conventional vessels are considered. The IEC 61162-460 standard [18] describes the security requirements of the maritime navigation and radio communication equipment and systems onboard, for conventional ships. To the best of our knowledge, no previous work has addressed the problem of identifying the security requirements of the cyber-physical systems of the C-ES by leveraging a systematic approach.

A multitude of security requirements engineering methods exists and several works have compared methods, tools, and frameworks for security requirements elicitation [19], [20], [21]. Most of the reviews analyze the pros and cons of the reviewed methods and conclude with a recommendation on their appropriateness. Several of these, e.g., [22], [23] recommend the Secure Tropos methodology [10] as enjoying many of the desirable characteristics. The methodology has been used to extract security and privacy requirements in several cases, including the industrial internet of things [24], [25]. In addition, a framework which combines EBIOS, Secure Tropos and PriS methods to extract security, privacy, and safety requirements for connected vehicles has been proposed in [26]. As privacy is not relevant to the CPS systems under study, because no personally identified data are involved with the operation of these systems, based on these findings, Secure Tropos was selected as the most appropriate methodology for the analysis of the complex C-ES ecosystem and for the elicitation of its security requirements.

The Maritime Architecture Framework (MAF) [11] is a domain specific architectural methodology designed to overcome the challenge to coordinate the development of new systems between technology issues, governance aspects and users between existing architectures in the maritime sector. The MAF is derived from the successfully established architecture model in the energy domain named Smart Grid Architecture Model (SGAM) [27]. The main element of the MAF is the multidimensional cube, that combines different viewpoints to provide a graphical representation of the underlying maritime domain and the examined system architecture. The cube captures three dimensions, namely interoperability; hierarchical; and topological.

III. Security requirements elicitation process

The proposed process of security requirements elicitation for the C-ES is based on and adapted from [28] and [25], and is depicted in Fig. 1. In the first stage, entitled "Early requirements", the C-ES ecosystem's actors, goals, assets, and resources are identified. The outcome of this phase is an actor diagram and a number of goal diagrams. In the next stage, entitled "Late requirements", the actor diagram of the early requirements is extended with the introduction of the system as an actor that has a number of dependencies with the rest of the actors. In fact, these dependencies will be the functional and non-functional requirements of the system. In the third stage, entitled "Security analysis", based on the system requirements and data and control flows among actors, a global architecture of the C-ES is defined, along with security constraints. The outcome of the overall process is the security requirements.

This process is implemented by leveraging the Secure Tropos methodology [10], initially designed as a security-aware software systems development methodology that combines requirements engineering concepts, such as "actor", "goal", "plan", together with security engineering concepts such as "threat", "security constraint" and "security mechanism". Different ecosystem components, dependencies, interdependencies, connections, and interconnections among systems can be visually represented through this method as well as security related arguments, such as security constraints, threats, vulnerabilities, and countermeasures. The application of the methodology is supported by the SecTro tool [28].

A. Environment analysis

The first step in the process is the analysis of the environment of the system under examination. To this end, we leverage the MAF [11]. This framework enables the structured

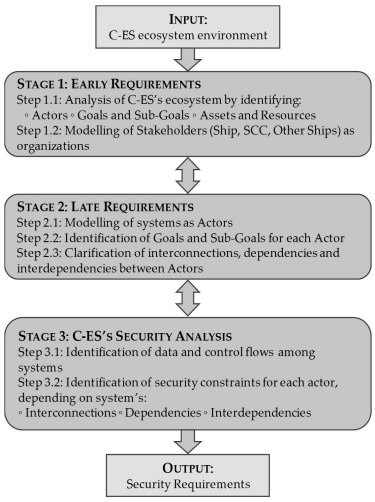


Fig. 1: Security Requirements Elicitation Process

representation of the maritime domain, in terms of elements of the ecosystem, such as information assets, people, and technology used. The environment is represented by means of the MAF multidimensional cube, where three layers, namely the C-ES, the SCC, and the communication link between them and the ecosystem's elements are depicted. Essentially, the environment of the C-ES is represented by the actors of a ship's ecosystem, goals, and dependencies among actors and goals.

Security requirements are most usefully defined as requirements for the operational and environmental constraints of the system under analysis [24]. Therefore, the detailed identification of such constraints is an important element of the security requirements elicitation process. The authors in [29] have already defined the operational constraints for the unmanned merchant ships [30] without, however, identifying constraints such as system vulnerabilities and potential cyber-attacks. The environmental constraints are inexorably linked to the C-ES's operational constraints, as they restrict the various goals and plans the ship has and can be exploited by adversaries, thus raising security issues. As the SCC is also a crucial part of the C-ES's ecosystem, environmental constraints for the SCC should also be identified. The identified environmental constraints for the C-ES are depicted and shortly described in Table I, and those for the SCC in Table II.

B. Organizational analysis

Stage 1 and Stage 2 of the security requirements elicitation process together constitute the organizational analysis of the ecosystem and of its elements. This is carried out by following steps 1.1 through 2.3 as indicated in Fig. 1. The analysis is carried out both for the ecosystem as a whole and for each one of the individual systems considered, namely the AIS, the ECDIS, and the GMDSS.

1) Ecosystem organizational analysis

Fig. 2 depicts the organizational view of the C-ES ecosystem where three entities have been identified: the Ship, the SCC, and other ships. Following the Secure Tropos methodology, these entities are represented as distinct organizations, by rectangles. Within the Ship, the bridge and the engine systems have been identified as actors, and are represented by circles; these interact with the external actors, such as the Human Machine Interface (HMI) of the SCC and other ships in the vicinity. Actors' boundaries are represented by dashed rounded rectangles that contain the goals and the sub-goals that the actors have to fulfill (represented by rounded rectangles), as well as the resources they require in order to satisfy those goals (represented by rectangles). The actors are defined based on their dependencies and interdepencencies, as depicted in Fig. 2. It should be noted that the organizational view of the ecosystem includes different types of data, depending on the actors these data derive from. For example, bridge systems communicate navigation, voyage and safety related data, while engine systems exchange engine related data.

2) AIS organizational analysis

The AIS provides information intended to facilitate the monitoring of traffic, thus contributing to ensuring the ship's safety and to increasing the situational awareness. The AIS exchanges data with six different navigational subsystems and two external actors, namely the SCC and other ships in the vicinity. The transmitted data can be static, voyage, dynamic, and safety-related, depending on the system interconnections and interdependencies, as it is depicted in the full organizational view[1] of the AIS.

3) ECDIS organizational analysis

The ECDIS provides and transmits information regarding the ship's voyage. Its full organizational view[2] includes eight internal and two external actors. The internal actors are the sub systems of the Navigation system and the external actors are the SCC and the ship controller. It is worth noting that although the ship controller is an on board system, it has been characterized as an external actor because it is not a sub-system of the navigation system. The goals and the sub goals of each actor have been identified taking into account the corresponding resources, i.e. the exchange data among actors; these can be static, dynamic, voyage and safety-related data.

4) GMDSS organizational analysis

The GMDSS ensures the rapid alerting of (no)shore authorities in the event of emergency. Its organizational view[3] includes the ship controller's sub systems, characterized as the internal actors. The external actors are the on board systems and sub systems which GMDSS interacts with, and the SCC. The goals and sub goals of each actor have been defined considering the type of the signals and data that are transmitted; these indicate dependencies and interdependencies. Transmitted signals and data are the resources that are required for each actor to accomplish its goals. The GMDSS is interdependent with

[1] https://drive.google.com/open?id=1uzLTvcqeGcVDS6BT4n8Oh7IwCcDGgNLE
[2] https://drive.google.com/open?id=1bw3vvi1UseVI40TnVo0TwXRxz1Q_pYjG

Constraint	Short description
Weather conditions	Heavy weather conditions, such as strong winds and heavy mist where the visibility is limited.
Legal	Sail in congested waters, such as ports using specific legal framework or SCC's directions.
Communication	Support a multitude of communication technologies.
Geographic	Islands, reefs, mountains which may influence the ship's operation, and protection of the sea life.
Cyber attacks	Since the C-ES is comprised of cyber-physical systems, the infrastructure may be exploited by physical/cyber attacks.
Traffic	Several other entities in the ship's vicinity, either physical or virtual.
Emergency	Search and rescue operations is compulsory according to International Maritime Organization (IMO) guidelines.
Restricted areas	Operating in Special Emission Control Area (SECA); ship reporting area or other restricted areas.
Harbors	Navigation in different harbors which are characterized by different architectures, port authorities, and legal frameworks.
Human factors	Ensure the safety of people and handle unpredictable incidents which derive from them.
Port systems	The interaction with the automated port systems is continuous and crucial for the security and monitoring of the cargo.

TABLE I: C-ES environmental constraints

Contraint	Short description
Weather conditions	Harsh weather may cause malfunction to the external sensors or antennas of the SCC building and could affect the delay and latency of communication.
Legal	The SCC should follow the International maritime legislation and standards for the safe and secure ship's operation.
Communication	Loss of communication link or malfunction of the satellite provider may cause disruptions to the C-ES.
Geographic	The location of the SCC is essential for its smooth communication with both the vessel and the shipping company.
Cyber attacks	The SCC is comprised of cyber and physical systems, like the C-ES.
Natural disaster	Flood, fire or earthquakes may influence the environment of the SCC and its operation.
Different vendors	SCC systems developed by different vendors could cause interoperability issues.
Personnel	The environment of the SCC may be affected in case of a personnel leaving or dismissal.
Multi role environment	The SCC is an environment where humans with diverse professional expertise and roles co-exist and co-operate.
Port Authorities	The SCC should be able to effectively communicate and interact with port authorities.
Stakeholders	The SCC communicates and interacts with stakeholders in order to ensure the vessel's operations.

TABLE II: SCC environmental constraints

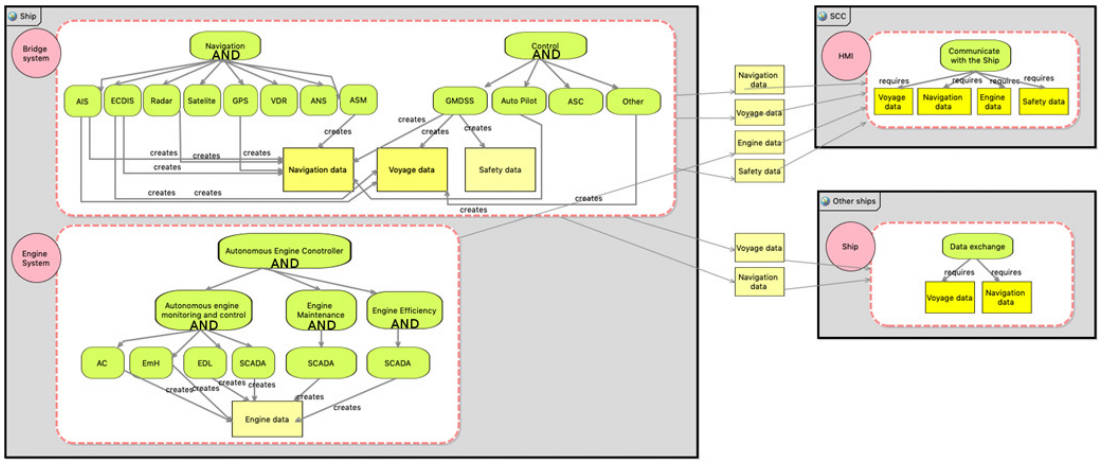


Fig. 2: General Ecosystem Representation

the onboard and onshore systems, the engine and navigation systems, and the SCC.

C. Security requirements

The organizational view of the ecosystem, as depicted in Fig. 2 constitutes the C-ES system's general architecture. Based on the functionality and the technical characteristics of the systems under study, the data and control flows are identified, as required in Step 3.1 of the security requirements elicitation process. These are depicted in Figs. 3, 4 and 6.

Step 3.2 requires the identification of the security constraints for each actor. In our case, these constraints are the elements of the *Parkerian Hexad*, i.e. *Confidentiality* – defined as "Limited observation and disclosure of knowledge"; *Integrity* – defined as "Completeness, wholeness, and readability of information and quality being unchanged from a previous state"; *Availability* – defined as "Usability of information for a purpose"; *Posession* – defined as "Holding, controlling, and having the ability to use information"; *Authenticity* – defined as "Validity, conformance, and genuineness of information"; and *Utility* – defined as "Usefulness of information for a purpose" [31].

[3]https://drive.google.com/open?id=1pQGSxM57s13GQkTrk3KhH7RPEhuARqz1

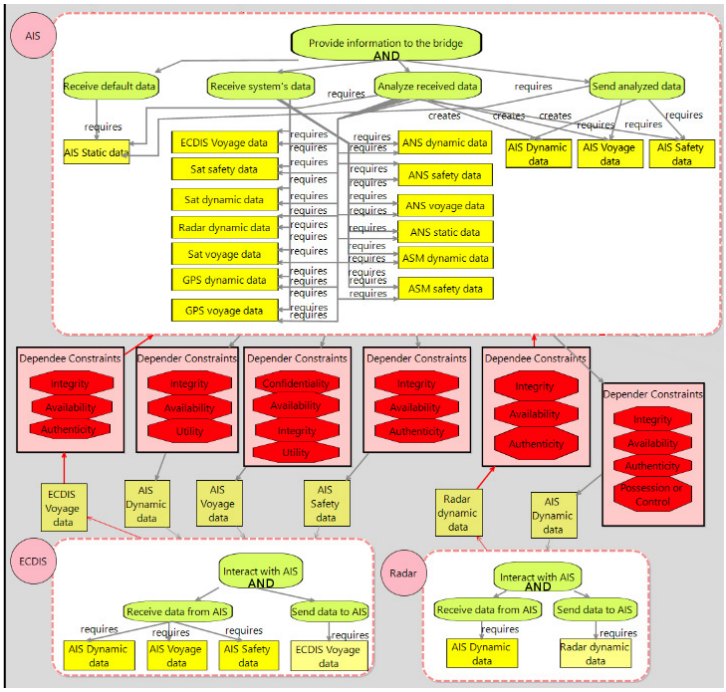


Fig. 3: AIS Security Requirements

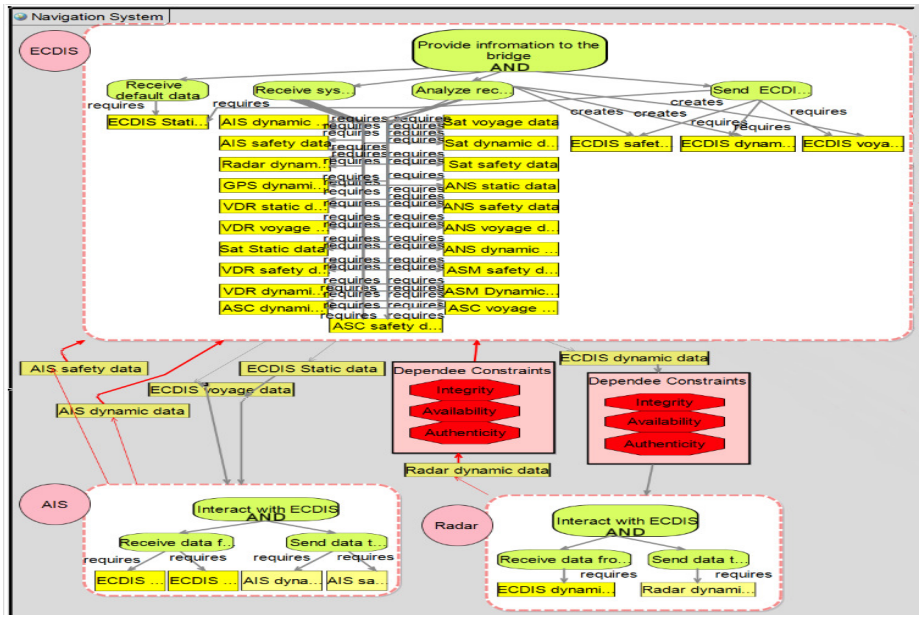


Fig. 4: ECDIS Security Requirements

According to [28], when using the Secure Tropos methodology, the security constraints in the system's goal diagram are the security requirements of the targeted system. The identified system functional and operational requirements lead to identifying the system goals, as well as the processes and resources utilized to achieve the identified goals. The security constraints which will protect the identified processes and goals are identified by considering the Parkerian hexad. An example of this procedure follows: Two identified security requirements are: "The connectivity between system and external actors and between on board systems must be continuous" and "Voyage related data transmitted to the SCC must be protected against tampering or damage". Following the Secure Tropos method, we analyze the environment of the targeted system and we identify its operational and functional requirements which are "Inform SCC about vessel's speed and position" and "Send Voyage data to SCC", respectively. Then, the goals and sub-goals that need to be achieved so as the system fulfills these requirements are identified. These are "1) Receive and analyze voyage data from ECDIS and Radar, and 2) Send analyzed data to SCC". The resources to achieve these goals are the AIS Voyage data. In order to design the system-to-be (in this case, a secure AIS system), the security constraints are identified. In this case, Availability and Integrity are identified as security constraints of the interconnections and interdependencies between the AIS and the SCC. Since a security requirement is the security constraint in the system's goal diagram, the resulting security requirements are: "The availability of the transmitted data between AIS and SCC should be ensured" and "The integrity of the processed and transmitted data must be protected". Considering the operational and functional requirements of the targeted system, and the potential threats to the AIS (Denial of Service, Tampering) [4] that could violate the identified constraints (Availability, Integrity) a system-specific security requirement is "Voyage related data transmitted to the SCC must be protected against tampering or damage". Since the protection of availability of the transmitted data is a common requirement for the three targeted systems, the availability requirement is refined to "The connectivity between system and external actors and between on board systems must be continuous" in the first group of requirements ("Common Security Requirements"). This flow of reasoning is depicted in Fig. 5.

The outcome of Stage 3 of the security requirements elicitation process, guided and supported by the SecTro tool, is the security requirements. These are presented in the sequel, following the classification scheme in the ISO 27001:2013 [32] and ISO 27002:2013 [33] standards. Several standards on the security of cyber physical systems are discussed in [34]. These include the ISO 27k family; NEC's CIP family of standards; and the ISA IEC IEC-62443 series. Also relevant are standards on software security requirements (such as e.g., ECSS-Q-ST-80 C, IEEE 830-1998, ISO/IEC 25010, ISO/IEC 27034-1, and ISO/IEC 27034-3). In the maritime domain, [17] provides a classification of cyber security requirements. As the ultimate goal of this research is to propose cyber security requirements for the whole C-ES ecosystem, we have decided

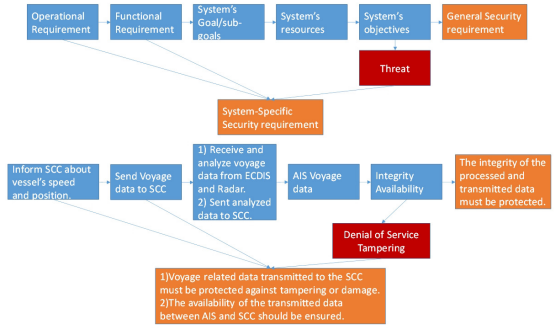


Fig. 5: Security Requirements Elicitation Process

to use the ISO 27001-27002 standards for **presenting** the requirements, as these pertain to organizations rather than isolated systems, be they software or otherwise. This will greatly facilitate their integration with additional requirements derived from other elements of the C-ES ecosystem. Using the classification in [17] could have been an alternative; however, we opted for a de jure standard rather than an industry proposal. Two groups of requirements are presented: common and system-specific. The former group includes requirements applicable to all three studied systems, whereas the latter includes requirements pertinent to each individual system.

1) Common Security Requirements

Human resource security: i) The system administrator must be well trained and aware of system functional and non-functional requirements (e.g., AIS modes and communication capabilities). **Asset management**: i) Data and signals must be identified and classified into protection levels; ii) A documentation of third-party components, versioning, and published system vulnerabilities must be maintained; **Access control**: i) A strong password policy must be enforced which will specify the length and the lifetime of each combination of the credentials (e.g., Passwords to log in to the ECDIS should be reguralry changed); ii) The non-repudiation and traceability of actions performed either from the SCC or physically to the on board system must be ensured with appropriate authentication mechanisms; iii) The system must be able to implement lock mechanisms when requested by the system administrator or after a configurable time of idleness; iv) The number of consecutive login attempts to the system must be specified; v) The system must support multi-factor authentication; vi) The system must accept inputs only from authorized entities, by authorized maritime actors. **Cryptography**: i) The system must support encryption algorithms able to promote data confidentiality and integrity, and to satisfy data transmission timing requirements during the voyage; ii) Data transmitted to external and internal actors should be encrypted by using appropriate -in each case- cryptographic mechanisms (e.g., Dynamic data sent from ECDIS to Radar, Global Positioning System (GPS), and Advanced Sensor Module (ASM) systems must be encrypted); iii) Stored data should be appropriately encrypted, the strength of the encryption mechanism depending on their type and the possible pertinence of maritime legal

or regulatory requirements. **Physical and Environmental Security**: i) The physical integrity of the on board or SCC sensors must be protected; ii) The system must be installed so as to prevent physical damages, such as flooding or fire; iii) All physical and virtual connection points of the system must be appropriately protected or blocked (e.g., USB ports or any other Human interface device-HID). **Operations security**: i) Both on board and SCC systems must be able to operate under network stress situations such as a Denial of Service attack; ii) Security mechanisms must be implemented in order to protect the system from malicious code; iii) Frequent system data backup should be maintained (e.g., ECDIS voyage data should be backed up regularly to the VDR); iv) The system must be able to determine whether an action taken has been performed by a system on board or by a human user remotely from the SCC; v) The integrity of the static, processed, and transmitted data must be protected; vi) The confidentiality of data in transit and in storage must be protected; vii) The freshness of data should be ensured; vii) The authenticity of services, transmitted data, and software sources must be ensured (e.g., AIS updates or ECDIS charts updates should be performed by authorized sources/vendors); viii) The utility of the dynamic and voyage data should be ensured; and ix) The measures to protect the confidentiality and integrity of data should not downgrade their utility. **Communication Security**: i) The confidentiality and integrity of the data exchanged between internal (on board systems and external actors (SCC or other vessel) should be ensured by appropriate mechanisms depending on the actors and the type of the data in transit; ii) The segregation of the on board components in different trust levels must be ensured; iii) The connectivity between system and external actors and between on board systems must be continuous; iv) On board systems must be mutually authenticated; v) The traffic from and to the system must be monitored, vi) The systems should be able to control the sent data considering the actor and the type of the data in transit, vii) All external actors of the C-ES ecosystem must be able to determine the source of data flows originating from the onboard systems; viii) The data exchange between on board systems should be established in a way such that their authenticity can be verified, ix) The systems must use transport layer security to protect the data in transit, x) The system should support mechanisms to detect rogue data packets, xi) The services between on board systems and external actors (SCC/other vessel) must be authenticated; xii) There should be redundancy of communication channels between on board) systems; xiii) The maximum allowable latency in system-to-system communication should conform to pertinent standards and to the systems' operational requirements. **System acquisition, development and maintenance**: i) System development and deployment must be performed following pertinent cyber-security standards; ii) The update process must be protected against time-of-check vs time-of-use attacks; iii) The source of the software must be authenticated; iv) Both on board and shore based systems must be maintained regularly; v) The system should be properly installed, taking into account network segmentation and physical access; vi) System updates/upgrades must be performed only by authorized entities; vii) The integrity of the maintenance process must be ensured to prevent malicious intrusions, viii) System maintenance must be performed only by well trained personnel; ix) The configuration and installation of the system must be performed by authorized personnel; x) The vessel's infrastructure must be well designed and the corresponding systems appropriately installed according to on the type of the ship; and xi) The system must not allow downgrading to old system software versions. **Supplier relationships**: i) Appropriate mechanisms must be employed to validate hardware, software, and data from the suppliers; and ii) Strict review of the security policies of the system's vendor must be undertaken. **Information security incident management**: i) The system must detect and produce an alert on abnormal numbers of requests, such as by a user or an external actor; ii) The system's functional and non-functional requirements should be maintained during a security incident such as e.g., GMDSS signal jamming; and iii) The SCC must be notified when a system anomaly has been detected. **Information security aspects of business continuity management**: i) The continuity of system operations must be ensured; ii) The system on board or on shore must be able to operate using alternative power sources; iii) The system must be able to operate 24/7; and iv) Redundant systems should be installed taking into account the operational complexity [4]of the C-ES and the system operations. **Compliance**: i) Formal certification of compliance with the pertinent legislative and regulatory requirements must be obtained.

2) AIS-specific Security Requirements

A part of the security requirements view of the AIS is depicted in Fig. 3. The full requirements view[5]is omitted in the interest of saving space.

Operations security: i) The AIS should implement the security services in order to protect the system from loss of control or possession of information; and ii) Voyage data, such as destination port or cargo related information should be confidential to prevent potential leakage to adversaries. **Communications security**: i) The communication channel with the radar system should be redundant; and ii) Voyage related data transmitted to the SCC must be protected against tampering or damage. **Access control**: i) Reliable authentication mechanisms must be in place in order to uniquely identify the actors reading, modifying, and transmitting AIS data, as well as to authenticate the system itself and its services; and ii) The AIS must be able to implement lock mechanisms (e.g., lock HMI screen) upon request by the administrator or after a configurable time of idleness. **Cryptography**: i) The authenticity of AIS functions (e.g., request, read, process, and send) must be ensured by using security techniques such as digital signatures.

3) ECDIS-specific Security Requirements

A part of the security requirements view of the ECDIS is depicted in Fig. 4. The full requirements view [6]is omitted in

[4]The C-ES's operational complexity depends on the mission and the environment of the vessel, as well as on its level of autonomy.

[5]https://drive.google.com/open?id=127DIgy9QR4H1b5K3-40Kx3KfDVyYsGEy

the interest of saving space.

Human resource security: i) The ECDIS administrator must be trained and able to distinguish rogue data packets. **Access control**: i) The use of ECDIS must be restricted only to authorized and well trained personnel. **Communication Security**: i) The ECDIS must be able to control the flows of voyage-related data sent to other ships and to the SCC; ii) The ECDIS should be able to audit sent and received data to external actors; iii) Safety-related information transmitted by the ECDIS must be authenticated; and iv) The communication between the ECDIS and the satellite system should be continuously available.

4) GMDSS-specific Security Requirements

A part of the security requirements view of the GMDSS is depicted in Fig. 6. The full requirements view [7]is omitted in the interest of saving space.

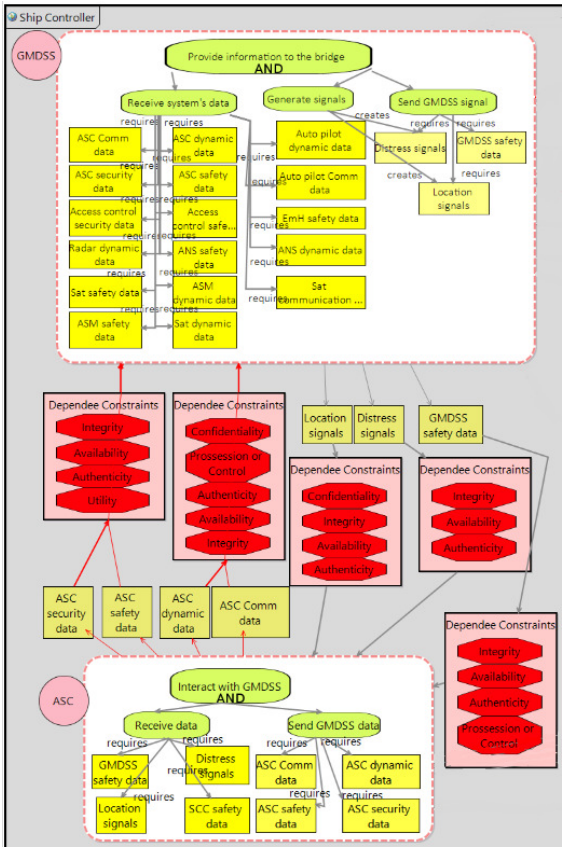


Fig. 6: GMDSS Security Requirements

Information security policies: i) A policy for installing the GMDSS components in the vessel's network should exist. **Access control**: i) The authenticity of the transmitted GMDSS signals and data in transit to the Autonomous Ship Controller (ASC), to other subsystems, and to the SCC must be ensured; and ii) Distress signals transmitted through the GMDSS must be verified by external actors such as SCC and other ship's subsystems such as the Autonomous Engine Monitoring and Control (AEMC) and Navigation systems. **Operations security**: i) The ASC must be able to provide security, safety, and dynamic data to the GMDSS, when needed. **Communication Security**: i) Safety signals transmitted through the GMDSS to other on board systems and external actors must be continuously available; ii) The GMDSS must be able to detect whether the signal/data comes from a legitimate user/system or from a malicious user; ii) The signals transmitted to external actors or subsystems must be appropriately encrypted. **System acquisition, development and maintenance**: i) GMDSS antennas must be appropriately installed.

[6]https://drive.google.com/open?id=1VljM1uibsuT--u7DuiIcPGnq5Y8TgSps

[7]https://drive.google.com/open?id=1errDRGKchm9UOZ_R0UCR-IRIbRrmAL9F

IV. Conclusions

We proposed a process for eliciting the security requirements of the C-ES ecosystem, based on the SecureTropos methodology and leveraging the Maritime Architecture Framework reference architecture as instantiated in the case of the C-ES. By applying the proposed process, we identified the security requirements for the three most vulnerable C-ES systems, namely the AIS, the ECDIS and the GMDSS. As future work we intend to address the issue of systematically deriving the security requirements of the C-ES viewed as a system-of-systems utilizing the requirements of each individual constituent system. Additionally, we intend to extend our work by allowing the combined elicitation of security and safety requirements.

References

[1] SINTEF, Shipping 4.0 presented at Singapore Maritime Week. [Online]. Available: https://www.sintef.no/en/latest-news/shipping-4.0-presented-at-singapore-maritime-week/

[2] J. Cross and G. Meadow, "Autonomous ships 101," *Journal of Ocean Technology*, vol. 12, pp. 23–27, 2017.

[3] K. Tam and K. Jones, "Cyber-risk assessment for autonomous ships," in *International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*. IEEE, 2018, pp.1–8.

[4] G. Kavallieratos, S. Katsikas and V. Gkioulos, "Cyber-Attacks Against the Autonomous Ship," in *SECPRE 2018, CyberICPS 2018. Lecture Notes in Computer Science, vol 11387*. Springer, 2018, pp.20–36.

[5] USCG, Cyber Incident Exposes Potential Vulnerabilities Onboard Commercial Vessels. [Online]. Available: https://www.dco.uscg.mil/Portals/9/DCO%20Documents/5p/CG-5PC/INV/Alerts/0619.pdf

[6] M. Jones, Spoofing in the black sea: What really happened?. [Online]. Available: https://www.gpsworld.com/spoofing-in-the-black-sea-what-really-happened/

[7] MARAD, 2019-012-persian gulf, strait of hormuz, gulf of oman, arabian sea, red sea-threats to commercial vessels by iran and its proxies. [Online]. Available: https://www.maritime.dot.gov/content/2019-012-persian-gulf-strait-hormuz-gulf-oman-arabian-sea-red-sea-threats-commercial-vessels

[8] G. Kessler and JP Craiger and JC Haass, "A Taxonomy Framework for Maritime Cybersecurity: A Demonstration Using the Automatic Identification System," *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 12, no. 3, p. 429, 2018.

[9] S. Katsikas, "Cyber Security of the Autonomous Ship," in *Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security*. ACM, 2017, pp.55–56.

[10] H. Mouratidis and P. Giorgini, "Secure tropos: a security-oriented extension of the tropos methodology," *International Journal of Software Engineering and Knowledge Engineering*, vol. 17, no. 2, pp. 285–309, 2007.

[11] B. Weinert, A. Hahn and O. Norkus, "A domain-specific architecture framework for the maritime domain," *in Informatik 2016*, H.C. Mayr and M. Pinzger, Eds, Gesellschaft für Informatik e.V., 2016, pp. 773–784.

[12] L. Kretschmann, Ø. J. Rødseth, B. S. Fuller, H. Noble, J. Horahan and H. McDowell, "MUNIN D9.3: Quantitative assessment," p. 150, 2015, project co-funded by the European Commission.

[13] Body of Knowledge and Curriculum to Advance Systems Engineering Editorial Board, "The Guide to the Systems Engineering Body of Knowledge (SEBoK), v. 2.0," [Online]. Available: www.sebokwiki.org.

[14] Ø. J. Rødseth, B. Kvamstad, T. Porathe and H. Burmeister, "Communication architecture for an unmanned merchant ship," in *Proceedings of MTS/IEEE OCEANS - Bergen*. pp.1–9, 2013.

[15] M. Höyhtyä, J. Huusko, M. Kiviranta, K. Solberg and J. Rokka, "Connectivity for autonomous ships: Architecture, use cases, and research challenges," in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2017, pp.345–350.

[16] Bureau Veritas, "Guidelines for Autonomous Shipping," Tech. Rep., 2017, [Online]. Available: https://www.bureauveritas.jp/news/pdf/641-NI_2017-12.pdf.

[17] DNVGL, "Cyber security capabilities of control system components," Tech. Rep, 2018.

[18] International Electrotechnical Commission- IEC, "Maritime navigation and radiocommunication equipment and systems," NEK IEC 61162-460:2018, p. 152, 2018.

[19] N. Mead, "How To Compare the Security Quality Requirements Engineering (SQUARE) Method with Other Methods," Tech. Rep., August 2017, [Online]. Available: https://apps.dtic.mil/dtic/tr/fulltext/u2/a471104.pdf.

[20] A. Nhlabatsi, B. Nuseibeh, and Y. Yu, "Security Requirements Engineering for Evolving Software Systems: A Survey," *International Journal of System of Systems Engineering*, vol. 1, pp. 54–73, 2010.

[21] A. Pattakou, C. Kalloniatis, and S. Gritzalis, "Security and privacy requirements engineering methods for traditional and cloud-based systems: A review," *Cloud Computing*, vol. 155, 2017.

[22] D. Mellado, C. Blanco, L. Sánchez, and E. Fernández-Medina, "A systematic review of security requirements engineering," *Computer Standards & Interfaces*, vol. 32, no. 4, pp. 153–165, 2010.

[23] D. Muñante, V. Chiprianov, L. Gallon, and P. Aniorté, "A review of security requirements engineering methods with respect to risk analysis and model-driven engineering," in *Proceedings of International Conference on Availability, Reliability, and Security*. Springer, pp.79–93, 2014.

[24] V. Diamantopoulou and H. Mouratidis, "Applying the physics of notation to the evaluation of a security and privacy requirements engineering methodology," *Information & Computer Security*, vol. 26, no. 4, pp. 382–400, 2018, [Online]. Available: https://doi.org/10.1108/ICS-12-2017-0087.

[25] H. Mouratidis and V. Diamantopoulou, "A Security Analysis Method for Industrial Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 4093-4100, 2018.

[26] C. Kalloniatis, V. Diamantopoulou, K. Kotis, C. Lyvas, K. Maliatsos, M. Gay, A. Kanatas, and C. Lambrinoudakis, "Towards the design of an assurance framework for increasing security and privacy in Connected Vehicles," *International Journal of Internet of Things and Cyber-Assurance*, 2019.

[27] CEN-CENELEC-ETSI Smart Grid Coordination Group, "Cen-cenelec-etsi smart grid coordination group smart grid reference architecture," p. 107, 2012.

[28] M. Pavlidis and S. Islam and H. Mouratidis, "A CASE tool to support automated modelling and analysis of security requirements, based on secure tropos," in *International Conference on Advanced Information Systems Engineering*. Springer, 2011, pp.95–109.

[29] Ø. J. Rødseth and Å. Tjora, "A system architecture for an unmanned ship," in *Proceedings of the 13th International Conference on Computer and IT Applications in the Maritime Industries (COMPIT)g*. 2014, p. 13.

[30] "MUNIN – Maritime Unmanned Navigation through Intelligence in Networks," [Online]. Available: http://www.unmanned-ship.org/munin/.

[31] D. B. Parker, "Toward a New Framework for Information Security?," *Computer Security Handbook, Sixth Edition*, S. B. Michel and E. K. E. Whyne, Eds. Wiley, 2012, ch. 3, pp. 3.1–3.23.

[32] International Organization for Standardization, ISO, "ISO/IEC 27001:2013 Information technology — Security techniques — Information security management systems — Requirements," 2013.

[33] International Organization for Standardization, ISO, "ISO/IEC 27002:2013 Information technology Security techniques Code of practice for information security controls," 2013.

[34] S. Ali, A. T. Balushi, Z. Nadir, and O. K. Hussain, "Standards for CPS," in *Cyber Security for Cyber Physical Systems*, Springer, 2018, pp. 161–174.

Georgios Kavallieratos Georgios Kavallieratos received the B.Sc. degree in computer science and the M.Sc. degree in digital systems security in 2016 and 2018, respectively, from the University of Piraeus, Piraeus, Greece. He is currently working toward the Ph.D. degree in security of the cyber-enabled ship in the Dept. of Information Security and Communication Technology, at the Norwegian University of Science and Technology. His area of expertise focuses on cyber-physical systems security and maritime cyber-security.

Vasiliki Diamantopoulou Vasiliki Diamantopoulou is an Adjunct Professor at the Department of Information and Communication Systems Engineering at the University of the Aegean and a Senior Researcher at the Department of Digital Systems at the University of Piraeus, Greece. She has worked as a Research Fellow at the School of Computing, Engineering and Mathematics at the University of Brighton, UK. The focus of her publications is on Privacy and Security of Information Systems, eGovernment and Interoperability Frameworks, and eBusiness and Innovation of Information Systems.

Sokratis K. Katsikas Sokratis K. Katsikas is the Rector of the Open University of Cyprus, and Professor with the Center for Cyber and Information Security, Norwegian University of Science and Technology. His research interests lie in the area of information and communication systems security. He has authored or co-authored more than 280 journal publications, book chapters and conference proceedings publications and he has participated in more than 70 funded national and international RD projects in these areas. He is serving on the editorial board of several scientific journals, he has authored/edited 41 books and has served on/chaired the technical programme committee of more than 680 international scientific conferences.

8 Article IV: Cybersecurity and Safety Co-Engineering of Cyberphysical Systems - A Comprehensive Survey [4]

future internet

MDPI

Article

Cybersecurity and Safety Co-Engineering of Cyberphysical Systems—A Comprehensive Survey

Georgios Kavallieratos [1,*], Sokratis Katsikas [1,2], Vasileios Gkioulos [1]

[1] Deparetment of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik 2815, Norway, georgios.kavallieratos@ntnu.no (G.K.); sokratis.katsikas@ntnu.no (S.K.); vasileios.gkioulos@ntnu.no (V.G.)

[2] Faculty of Pure and Applied Sciences, Open University of Cyprus, Latsia 2220, Cyprus; sokratis.katsikas@ouc.ac.cy

* Correspondence: georgios.kavallieratos@ntnu.no

Received: 28 March 2020; Accepted: 8 April 2020; Published: 11 April 2020

Abstract: Safeguarding both safety and cybersecurity is paramount to the smooth and trustworthy operation of contemporary cyber physical systems, many of which support critical functions and services. As safety and security have been known to be interdependent, they need to be jointly considered in such systems. As a result, various approaches have been proposed to address safety and cybersecurity co-engineering in cyber physical systems. This paper provides a comprehensive survey of safety and cybersecurity co-engineering methods, and discusses relevant open issues and research challenges. Despite the extent of the existing literature, several aspects of the subject still remain to be fully addressed.

Keywords: safety; cybersecurity; co-engineering; cyber physical systems

1. Introduction

The unification of embedded systems with communication technologies gave rise to "Cyber Physical Systems (CPS)". Such systems are deployed in several application domains, such as automotive, smart manufacturing, and healthcare. Due to the "double nature" of such systems—merging of the cyber and the physical worlds—ensuring both safety and cybersecurity are important prerequisites to their reliable operation.

Thus, the study of potential hazards and threats, and the assessment of safety and cybersecurity risks that potential accidents and cyberattacks pose, is important; it is also, usually, complicated. This is because there exist strong dependencies between the two domains, even though cases where they are independent do also exist. Three types of such dependencies have been identified and studied in Reference [1]:

1. **Conditional dependencies:** Safe operations may be conditioned by cybersecurity, for example, malicious modifications of sensor data or control programs may prevent safety systems from protecting an installation in accidental conditions. Conversely, safety may be a condition for cybersecurity, for example, when unmanaged catastrophic conditions weaken the security posture of a system or an organization, and lead to opportunistic malicious acts.
2. **Reinforcement:** Safety and cybersecurity measures can be complementary, for example, event and activity logging may be used both for attack detection and accident anticipation, as well as post-event analysis.
3. **Conflict:** If safety and cybersecurity are considered separately for the same system, it is possible that conflicting requirements or measures may be identified, for example, a safety requirement for an automatic door shutting system, would be to leave the door open, whereas a security requirement would be to leave the door locked in case of failure.

Therefore, there is a need to analyze both the safety and the cybersecurity of CPSs by employing a single approach. Such an approach can be employed to identify system faults/vulnerabilities, hazards/threats, safety/security requirements, and to assess safety/security risks.

Security engineering approaches aim to identify, assess, and manage risks related to the confidentiality, integrity, and availability of the targeted system/component. Various methods for security engineering have been proposed in the literature, that focus on different phases of the system lifecycle. General approaches assess and manage the overall security risk of a system, whilst methods also exist that facilitate the analysis in a particular phase of the lifecycle such as the requirements engineering, the threat analysis, the vulnerability analysis, or the risk analysis phases.

Safety engineering approaches aim to identify, assess, and manage risks related to the safety of the system, of humans, and of the environment. Various safety engineering methods exist, designed for different application domains and for different types of system safety (e.g., functional safety, operational safety etc.). Similarly with the security engineering methods case, different approaches exist for hazard analysis, fault analysis, cause analysis, and safety risk analysis and management.

Accordingly, security and safety co-engineering approaches aim to identify, assess, and manage risks related to both security and safety in systems which are influenced from both the cyber and the physical world/environment. Such approaches are classified according to their goal in three categories [2]:

- **Security-informed safety approaches:** Approaches that extend the scope of safety engineering by adapting cybersecurity-related techniques.
- **Safety-informed security approaches:** Approaches that extend the scope of security engineering by adapting safety-related techniques.
- **Combined safety and security approaches:** Combined approaches for safety and cybersecurity co-engineering.

In recent research, many proposals for security and safety co-engineering methods have appeared, and some survey articles have reviewed these proposed methods in varied degrees of depth and scopes. Pièrre-Cambacédès et al. [3] surveyed the differences and similarities between safety and security aspects focusing on their dependencies per application domain. Kriaa et al. [4] conducted a survey of safety and security methods and analyzed methods for industrial control systems. Various safety and security risk assessment methods, categorized according to their application domain, were reviewed by Chockalingam et al. [5]. Abulamddi [6] surveyed existing methods for safety and security requirements engineering in CPSs. A systematic literature review was conducted by Lisova et al. [7] that focused on already developed and evaluated methods. Lyu et al. [8] provided a short survey, in which five integrated safety and security co-engineering methods were analyzed. Finally, Paul and Rioux [2] provided an extended bibliography of research papers on safety and cybersecurity co-engineering since the early 90s without, however, analyzing them.

In this paper we revisit previous surveys on cybersecurity and safety co-engineering approaches; we report on the results of a systematic literature survey of such approaches that have not been reviewed before; we define a multi-attribute taxonomy and we use it to analyze such approaches; and we discuss pertinent open issues and research challenges. The overall contribution of this paper is a comprehensive discussion on the recent advances in cybersecurity and safety co-engineering. A summary of the contributions of the paper follows:

- A comprehensive review of sixty eight methods for cybersecurity and safety co-engineering, of which nine have not been reviewed before.
- The development of a multi-attribute taxonomy of cybersecurity and safety co-engineering methods, encompassing inter alia all attributes used in previous surveys.
- A discussion on open issues, not fully addressed by existing approaches for cybersecurity and safety co-engineering, that give rise to research challenges.

The remainder of this work is structured as follows: Section 2 reviews related work; it is divided into two subsections, one on previous surveys and another on approaches not previously reviewed. In Section 3 we define a multi-attribute taxonomy of cybersecurity and safety co-engineering approaches and we employ it to analyze the reviewed approaches. In Section 4 we discuss the results of the analysis, and we identify issues not fully addressed by existing approaches that give rise to research challenges. Finally, in Section 5 we summarize our conclusions and we outline our future research plans.

2. Related Work

2.1. Previous Reviews

There are two types of reviews; *narrative* and *systematic* reviews [9]. *Narrative* reviews aim to identify studies in the literature that describe a specific problem of interest. In such reviews, systematic guidelines regarding the searching method, the identification of research questions, and the screening process are not considered. Thus, such reviews do not provide a comprehensive understanding of the stated problem.

Systematic reviews are methodical approaches to identify, analyze, and criticize the results concerning predefined research questions. Such reviews aim to provide inclusive results regarding a well predefined problem with specific research questions. Specific processes and guidelines exist for conducting a systematic literature review, including on how to formulate the research questions; research the literature; screen and select the results; and analyze and document the conclusions.

A total of six reviews of joint safety and security analysis methods have been identified in the recent literature. Of those five, References [3–6,8] are narrative and only one [7] is systematic. Even though the total number of methods reviewed therein is 60, surprisingly, the intersection of the set of methods reviewed in Reference [7] and of the set of methods reviewed in all other reviews counts only seven elements.

Pièvre-Cambacédès et al. [3] conducted a survey of various safety and security approaches and studied the potential adoption of a safety approach for security analysis and vice versa. Although this work provides insight into safety and security concepts by analyzing the relevant terminology, the methods that are surveyed are not combined approaches but traditional safety and security methods. Specifically, safety standards along with hazard and risk analysis methods on one hand, and vulnerability and threat analysis methodologies on the other have been studied. The surveyed methods have been classified according to type (*safety to security* or *security to safety*); according to the approach taken (*Architectural concepts, Graphical modeling, structured risk assessment, and Testing*); and according to safety characteristics (*fault prevention, fault tolerance, fault removal, and fault forecasting*).

Kriaa et al. [4] provide a survey of approaches combining safety and security risk analysis for industrial infrastructures. Thirty nine methods were analyzed and grouped considering various criteria: the way each method treats safety and security (*unification/integration or harmonization*); the lifecycle phase (*operational/requirements or design*) in which the studied system is; and the type of the risk assessment approach (*quantitative/qualitative*). The methods are classified into *generic* and *model-based*. Methods in the former group describe the lifecycle stages and the sequence of activities in each stage, whereas the latter includes *graphical or non-graphical* methods, that may be supported by software tools. Furthermore, an overview of the safety and security standards for industrial infrastructures is presented. Finally, by analyzing the safety and security dependencies and interdependencies, the authors concluded that safety and security are complementary and should be treated jointly to improve the risk assessment process.

Chockalingam et al. [5] surveyed several integrated safety and security risk assessment methods and identified their key characteristics. The analysis was performed considering five criteria: First the identified approaches were classified according to the number of the *citations* that they had received in the scientific literature. The authors argue that the research community started to recognize the

importance of the combination of safety and security analyses in 2014 and 2015, with the most prominent methods being the Extended Fault Tree (EFT) and the Extended Component Fault Tree (E-CFT). Additionally, the approaches were grouped according to the steps involved in the risk assessment process. Specifically, the authors identified two types of integrated methods: the *sequential* integrated safety and security risk assessment method, and the *non-sequential* method. The third criterion is the stages of the risk management process (*risk identification, risk analysis, risk evaluation* that the method addresses. Further, the identified approaches are classified according to how the integration is achieved: *1) Combination of a conventional safety assessment method and of a variation of it to assess security; 2) Combination of a conventional security assessment method and of a variation of it to assess safety; 3) Combination of a conventional safety and a conventional security method;* and *4) Other - no conventional safety or security assessment method used*. Finally, the survey categorized the approaches considering the *application and the application domain*; four out of seven are methods targeting the transportation domain.

Abulamddi [6] surveyed integrated techniques for requirements engineering in CPSs; eight methods focusing on the requirements engineering phase of the lifecycle were identified and analyzed. The techniques were classified as *safety and security requirements* or *accident analysis*.

Lisova et al. [7] conducted a systematic literature review of joint safety and security analysis methods. The search was performed using the keywords ("safety" AND "security" AND "analysis") in four scientific databases (IEEE, ACM, Web of Science, and Springer link). Thirty three methods that analyze safety and security of CPSs early in the development phase have been identified. Five characteristics of these methods were considered: *application domain; stage in the system lifecycle; association with relevant standards; existence of validation step; and origin of contribution*. Additionally, the identified methods were classified according to the relationship between safety and security in the analysis process (*Unified, Parallel*), and the overall goal of the analysis (*combined safety and security; security informed safety, safety informed security*). Additionally, the yearly distribution of the reviewed papers, based on their security and safety focus was studied.

Finally, Lyu et al. [8] surveyed ten methods for safety and security analysis; these included five integrated approaches. The identified approaches were compared considering characteristics such as: *quantitative/qualitative, model-based/system-based, top-down/bottom-up analysis, and hierarchical/dynamic analysis*. The authors identified the pros and cons of each method and the technical gaps of the interplay of safety and security. Most of the integrated approaches analyzed in this work take a qualitative risk management approach.

2.2. Methods not Included in Previous Reviews

2.2.1. Search Method

Additional methods for safety and cybersecurity co-engineering have been identified in the following research databases: ACM Digital Library, Science Direct, Scopus, and IEEE Xplore. The search process was carried out by searching with the groups of keywords; (**Safety** AND **Security** AND **Cyber-physical systems**) and (**Safety** AND **Cybersecurity** AND **Cyber-physical systems**). The initial search returned 1313 results. The selection of the articles to be considered was performed according to the criteria listed below:

- The article must be explicitly related to a cybersecurity and safety **co-engineering** methodology.
- The article must not be included in previously published reviews.

This process resulted in the methods reviewed in the next section.

2.2.2. Methods

US^2 [10]: This is a unified approach that analyzes safety hazards and security threats for CPSs in automotive vehicles, by leveraging a simple quantitative scheme. It aims to analyze safety and

security concepts simultaneously, and to obtain consistent safety and security requirements and countermeasures. The elicitation of requirements is based on the ISO 26262 Automotive Safety Integrity Level (ASIL) metric and on the Security Level (SEL) metric, proposed in this work. The analysis is initiated by identifying security threats; in the sequel whether these threats may inflict safety hazards is examined. If so, the ASIL is utilized to identify the corresponding safety and security requirements and countermeasures, otherwise the SEL is used.

STPA and Six Step Model [11]: It is an integrated approach to analyze safety and security issues and artefacts for autonomous vehicles. It is based on the SAE J3016, SAE J3061, and ISO 26262 standards. By leveraging the Six Step Model (SSM) [12], the authors integrated the System Theoretic Process Analysis (STPA) [13] and the ISO 26262 standard to enrich the SSM hierarchies and, particularly, the lists of functions, failures, and safety countermeasures. The method comprises six steps, similarly to the SSM. In Steps 1 and 2 the functions, structure, and processes of the CPSs in an autonomous vehicle are identified. Steps 3 and 4 pertain to the safety and security analysis of the targeted system. Namely, the Hazard Analysis and Risk Assessment (HARA) as defined in ISO 26262 [14] and STPA methods are followed, to identify hazards, failures, and requirements. The security analysis is based on TARA as defined by the SAE J3061 standard [15]. Finally, in steps 5 and 6 the safety and security countermeasures are identified, by analyzing the functional safety and security requirements, and added to the model. A software too to support the proposed methodology is under development.

FACT [16,17]: Failure-Attack-CounTermeasure (FACT) is a unified graphical approach for safety and security analysis. This approach facilitates the analysis of CPSs and can be used for verification, validation, monitoring, and periodic safety and security assessment. The proposed approach is based on the ISA84 [18] and ISA99 [19] standards. The integration of the two standards is achieved by merging the corresponding lifecycle phases, resulting in a unified lifecycle of fourteen phases. The FACT graph model is developed in phases 1-9 of the unified lifecycle. The graph is constructed by following four distinct steps: 1) Import failure trees; 2) Include safety countermeasures in the graph; 3) Import attack trees; and 4) Include security countermeasures. Further, by leveraging the FACT graph, the security and safety countermeasures can be mapped to the corresponding attacks and faults. This enables the identification of interrelated countermeasures and the analysis of their interdependencies. The proposed approach has been applied to analyze industrial control systems.

CRAF [20]: The Cyber Risk Assessment Framework (CRAF) aims to facilitate the safety and security analysis of a CPS during the whole system lifecycle. The main focus of the framework is to study how a loss of data security could have safety implications. The framework comprises three steps: 1) Communicating a decision; 2) Raising a conflict; and 3) Conflict resolution. CRAF treats safety and security separately, and utilizes traditional security and safety techniques and concepts (e.g., Threat analysis, Hazard analysis). CRAF aims to bridge the gap between safety and security by comparing and consolidating the security and safety data properties.

UFoI-E [21]: The Uncontrolled Flows of information and Energy (UFoI-E) method enables the analysis and representation of the dependencies of CPSs and facilitates their diagrammatic representation for risk analysis. It provides a generic CPS master diagram to distinguish the cyber and physical environments of the system under study. The method integrates the security and safety concepts from physical, control, and computer systems. The dependencies between information and control flows are studied to examine the causes the could provoke harm to humans, assets, or the environment. The method considers these cyber threats in the information domain that could provoke safety hazards in the energy domain. Security aspects are related to deliberate sources of risk while safety aspects are related to unintentional sources of risk. According to the UFoI-E, both an uncontrolled flow of information (security) and an uncontrolled flow of energy (safety) may result in physical harm.

AVES [22]: The Automated Vehicles Safety and Security Analysis Framework (AVES) aims to facilitate the safety and security analysis of autonomous vehicles by leveraging four relationship matrices and a Safety and Cybersecurity Deployment (SCSD) model. The first matrix describes the

hazards and the threats of the targeted vehicle along with the associated risks and the pertinent safety and security requirements. The second matrix describes the safety and security countermeasures, and the third analyzes the relationships among the countermeasures. The fourth matrix records and tracks the implementation status of the previously identified countermeasures. Finally, the fifth matrix incorporates the other four matrices into a meta-model to better analyze the system by leveraging various data from different matrices. The method is consistent with relevant safety and security standards, covering the vehicle's development lifecycle. AVES is implemented in eleven stages, that cover the concept; product development; production; operations; and service and decommissioning phases of the vehicle lifecycle, and is able to capture various aspects of the vehicle at different automation levels.

CPS master diagram [23]: The CPS master diagram is a hierarchical three-layer representation of the studied system in different process types. The lower level represents the system's physical layer, that consists of the energy flows and the physical interactions that control the flows. The middle level describes the real time information flows to control, and the third (top) level is the cyber layer which consists of information flows for monitoring and supervision. By leveraging the master diagram, experts from the safety or the security field may apply existing or new assessment approaches to analyze different CPS applications. The framework has been used for preliminary safety and security assessments in the maritime [23] and in the Internet of Things (IoT) [24] domains.

IoT medical devices [25]: This work proposes a method to analyze safety and security issues in IoT medical devices. The method is based on the STPA and an analyst is able to identify and assess accidents due to security threats that violate the functional safety. By leveraging this approach the analyst is able to analyze complex systems and perform a quantitative threat analysis by combining the EFT and the Defense Tree (DT) methods. The approach comprises seven steps: In step 1 the accidents, hazards, and safety constraints are identified; in step 2 the control structure of the system is constructed according to the STPA. In steps 3 and 4 the unsafe control actions are identified, as well as the hazards' causal factors, by employing the EFT and DT methods. In step 5 the probability of occurrence of the fundamental events of the EFT that was developed in the previous step is calculated; the estimate is based on both statistical data and the stakeholders' judgement. Finally, in step 6 the selection of the appropriate countermeasures is performed, by considering the impact of the identified accidents and the probability estimates of step 5. The proposed method has been applied to the case of an insulin pump for diabetic patients.

SARA [26]: SARA (Security Automotive Risk Analysis) provides a framework for facilitating threat modeling and the risk assessment processes for driverless vehicles. Although it is a security risk assessment method, it enables the analysis of safety issues inflicted by security threats. This is achieved by examining the impact on safety of the attack goal, and by estimating the safety severity and controllability metrics. SARA consists of four blocks: 1) Feature definition; 2) Threat specification; 3) Risk assessment; 4) Countermeasures. In the third block (risk assessment) security and safety experts identify attacks and the necessary metrics for the risk estimation, such as severity, observation, controllability, and highest attack likelihood. The proposed method was applied to the case of an autonomous car to present the potential impact of a malicious observer and of damaged road infrastructure on the vehicle.

3. Analysis

In order to provide a comprehensive description of the current landscape of methods for the joint analysis of safety and security of CPSs, the following list of attributes is used. Several of these attributes have been used in previous reviews or elsewhere. Each attribute is followed by a short description and a reference to the source(s) where it was originally used.

1. **Type of joint analysis (Type):** This attribute may assume either the value "Integrated (I)" or the value "Unified (U)". In the former case the analysis is done in two separate, yet interrelated

processes, whereas in the latter the analysis is performed following a single, unified process. The attribute was originally used in References [4,7].

2. **Model type (Model):** Describes the model that the analysis is based on. Possible values are "Graphical (G)", "Formal (F)" and "Both graphical and formal (Both)". In graphical methods the analysis is carried out by leveraging graphical models, whilst in formal methods the analysis is based on formulas, equations, and modelling languages. This attribute was originally used in References [4,8].
3. **Standards:** The method is informed by and leverages safety/security standards. Possible values are "Yes (Y)" and "No (N)". This attribute was originally used in Reference [7].
4. **Application domain (Domain):** The application domain(s) where the method is applicable or has been applied. Possible values are "CPS", "IoT", "Automotive (A)", "Control Systems (CS)" or combinations thereof. This attribute was originally used in References [5,7].
5. **Approach:** The type of approach followed. Possible values are "Quantitative (QNT)" and "Qualitative (QLT)". This attribute was originally used in References [4,8].
6. **Goal of the analysis (Goal):** Describes the overall goal of the analysis and whether the approach aims to ensure safety, security, or both. Possible values are "Security", "Safety" and "Both". This attribute was originally used in Reference [7].
7. **Lifecycle:** Describes in which phase of the system lifecycle the method is applied. Possible values are "Requirements (RE)", "Risk Analysis (RA)", "Any phase - Generic (GE)". This attribute was originally used in Reference [4].
8. **Stakeholders:** Describes which stakeholders are involved, either by applying it (users) or by giving input (participants); when applying the method. Possible values are "Safety experts (A)", Security experts (B)", "Developers (C)", "Designers (D)", and "Users or system experts (E)". This attribute was originally used in Reference [27].

The above attributes are depicted in Figure 1. Further, the following characteristics provide additional insight into understanding the operational capacity of each method; these have been inspired by the work in Reference [27]. Each of them, with the exception of *Process*, may assume the value "Yes (Y)", "No (N)", or "Partially (P)". *Process* may only assume the values "Yes (Y)" or "No (N)".

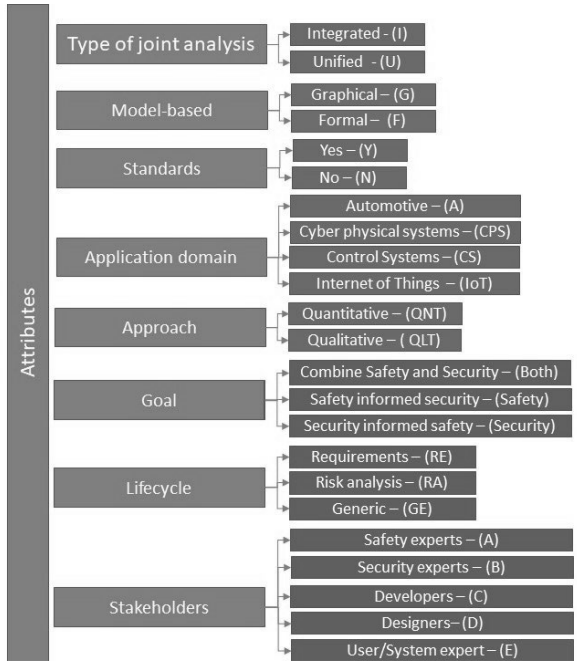


Figure 1. Attributes

1. **Process:** Is the method supported by a systematic and structured process?
2. **Scalability:** Does the method scale well with the size and complexity of the system under assessment?
3. **Creativity:** Does the method include mechanisms to stimulate creativity among the stakeholders? Examples of such mechanisms are guide-words, checklists and questionnaires.
4. **Communication:** Is the method offering features to facilitate communication between different stakeholders during its application? Examples of such features are guidelines, diagrams, schematics, and so forth.
5. **Conflict resolution (Conflict):** Does the method facilitate the identification and study of potential conflicts between safety and security aspects?
6. **Software tool (Tool):** Does a software tool or toolkit that supports the application of the method exist?

Method	Type	Model	Standards	Domain	Approach	Goal	Lifecycle	Stakeholders	Scalability	Creativity	Communication	Process	Conflict	Software
[10]	U	F	No	A	QLT	Both	RE	A/B	Y	Y	P	Y	N	N
[11]	I	G	Yes	A	QLT	Both	RA/RE	A/B/D	Y	Y	Y	Y	P	N
[16,17]	U	G	Yes	CS	QLT	Both	RA/RE	A/B/D	Y	Y	P	Y	Y	N
[20]	I	N/A	No	CPS	QLT	Security	RE/RA	A/C/E	P	P	Y	Y	Y	N
[21]	I	G	Yes	A	QLT	Safety	RA	A/B/C/E	Y	P	Y	Y	P	N
[22]	I	G	No	CPS	QLT	Both	RA/RE	A/B/C/D	P	Y	Y	Y	Y	N
[23]	I	G	No	CPS	QLT	Both	RA	A/B/D/E	P	Y	Y	Y	N	N
[25]	I	G	No	CPS	QLT	Safety	RA/RE	A/B/D	Y	Y	Y	Y	N	Y
[26]	I	N/A	No	CPS/A	QLT	Both	RA	A/B/E	N	P	N	Y	N	N
[28]	U	G	No	CPS	QLT	Safety	RE	A/B/C	N	Y	Y	Y	Y	N
[29]	I	N/A	Yes	CS	QLT/QNT	Safety	RA	A/B	P	Y	N	Y	N	N
[30]	U	G	Yes	CPS	QLT	Safety	RA	A/B	P	Y	P	Y	N	N
[31]	U	G	No	CPS	QLT/QNT	Both	RA/RE	A/B/D/E	Y	N	N	Y	P	P
[32]	U	F	Yes	CPS	QLT/QNT	Both	RA	A/B	P	Y	P	Y	N	N
[33]	U	F	Yes	CS	QLT/QNT	Both	RA/RE	A/B	N	P	P	Y	Y	N
[34]	U	Both	No	CS	QLT	Safety	GE/RE	A/B	P	Y	Y	Y	Y	N
[35]	U	Both	Yes	A	QLT	Safety	RE	A/B	Y	P	Y	Y	Y	N
[36]	U	Both	No	A	QLT/QNT	Safety	GE/RA	A/B/C	P	Y	Y	Y	Y	Y
[37]	I	G	No	CS	QLT/QNT	Both	RE	A/B/D	P	P	N	Y	N	N
[38]	U	F	No	CS	QLT/QNT	Both	RA	A/B/E	Y	P	P	Y	Y	Y
[39]	U	F	Yes	A	QNT	Safety	RE	A/B	P	P	Y	Y	N	N
[40]	U	F	Yes	A	QNT	Safety	RA	A/B/D/E	Y	Y	Y	Y	N	N
[41]	U	F	Yes	CPS	QLT/QNT	Safety	RA	A/B	N	Y	N	Y	N	N
[42]	U	G	No	A	QLT	Safety	RA	A/B/D	N	Y	Y	Y	N	N
[43]	I	Both	Yes	A	QLT	Both	GE/RE/RA	A/B/C/D/E	P	Y	Y	Y	N	Y
[44]	I	G	No	IoT	QLT/QNT	Safety	RA	A/B/C/D	Y	Y	Y	Y	Y	N
[45]	U	G	Yes	A	QLT	Both	RE	A/B	N	Y	P	Y	Y	X
[46]	U	G	Yes	A	QLT	Safety	RA/RE	A/B/C/D	Y	Y	P	Y	N	P
[47]	I	G	No	CPS	QLT	Safety	RE	A/B/D	P	Y	N	Y	N	N
[48]	U	N/A	Yes	A	QLT	Safety	GE/RA/RE	A/B/E	Y	Y	N	Y	N	N
[49]	U	G	No	CPS	QLT	Both	RA	A/B/C/D	Y	Y	Y	Y	N	P
[50]	U	F	No	CPS	QLT/QNT	Both	RA/RE	A/B/C	Y	Y	Y	Y	N	Y
[51]	U	Both	No	CPS	QLT/QNT	Both	RA	A/B	Y	N	P	Y	N	Y
[52]	I	N/A	Yes	CPS	QLT	Both	GE/RA/RE	A/B/D	Y	Y	Y	Y	Y	P
[53]	U	N/A	No	CPS	QLT	Both	GE/RA	A/B/D	Y	Y	Y	Y	N	P
[54]	U	N/A	No	IoT	QLT	Safety	GE/RA	A/B	Y	Y	Y	Y	N	P
[55]	I	G	Yes	CPS	QLT	Both	RA	A/B	Y	Y	Y	Y	N	N
[56]	U	G	Yes	CPS	QLT/QNT	Safety	RA	A/B	Y	Y	Y	Y	N	N
[57]	U	F	No	CS	QLT	Both	RE	A/B	Y	Y	Y	Y	N	P
[58]	U	N/A	No	N/A	QLT/QNT	Both	GE/RA	A/B	Y	N	N	Y	N	N
[59]	U	F	No	N/A	QLT/QNT	Both	GE/RE/RA	A/B	Y	Y	N	Y	N	N
[60]	U	N/A	Yes	CPS	QLT	Both	GE/RE/RA	A/B	Y	Y	P	Y	N	N
[61]	U	N/A	Yes	IoT	QLT	Safety	GE/RA/RE	A/B/C/D	Y	N	N	Y	N	N
[62]	I	N/A	No	CS	QLT	Both	GE/RA	A/B/C	N	P	P	Y	Y	N
[63]	I	N/A	No	CS	QLT	Safety	GE/RE	A/B/C/D	Y	Y	Y	Y	N	N
[64]	I	N/A	No	CS	QLT	Both	GE/RE	A/B	Y	Y	P	Y	Y	N
[65]	I	G	No	CS	QLT	Safety	RE/RA	A/B	Y	Y	N	Y	N	N
[66]	I	Both	No	CS	QLT/QNT	Safety	RA	A/B	P	Y	N	Y	N	N
[67]	I	Both	No	CPS	QLT/QNT	Safety	RE	A/B	Y	Y	P	Y	N	N
[68]	I	Both	No	CPS	QLT/QNT	Safety	RE/RA	A/B	P	Y	Y	Y	N	Y
[69]	I	Both	No	CS	QLT/QNT	Safety	RA	A/B	P	Y	P	Y	N	N
[70]	U	F	No	N/A	QLT/QNT	Security	RA/RE	A/B/E	P	Y	P	Y	N	Y
[71]	I	Both	No	CS	QLT/QNT	Both	RA	A/B	P	Y	P	Y	N	Y
[72]	I	G	No	CPS	QLT	Security	RE/RA	A/C/D	Y	Y	P	Y	N	N
[73]	I	Both	No	CS	QLT/QNT	Safety	RE	A/B	P	Y	Y	Y	N	Y
[74]	I	F	No	A	QLT/QNT	Security	RE/RA	B/D	Y	Y	N	Y	N	Y
[75]	U	Both	No	CS	QLT/QNT	Safety	RE/RA	B	P	Y	P	Y	N	Y
[76]	I	F	No	CPS	QLT/QNT	Both	RE/RA	A/B/D	Y	P	P	Y	N	Y
[77]	I	F	No	IoT	QLT	Both	RE	A/B	P	Y	P	Y	Y	Y
[78]	I	F	No	CPS	QLT/QNT	Security	GE/RE/RA	A/B	Y	Y	P	Y	N	N
[79]	I	F	No	CPS	QLT/QNT	Both	GE/RE	A/B	P	Y	P	Y	Y	N
[80]	I	F	No	A	QLT	Security	RE	B	P	Y	P	Y	N	Y
[81]	I	F	No	CPS	QLT	Both	RE	A/B/D	Y	Y	P	Y	N	Y
[82]	I	G	No	CPS	QLT	Safety	RE/RA	A/B/C/D	Y	Y	Y	Y	N	Y
[83]	I	G	No	A	QLT	Safety	RA/RE	A/B	Y	Y	P	Y	N	N
[84]	I	N/A	No	A	QLT	Safety	RA	A/B	P	Y	P	Y	Y	Y
[85,86]	I	G	No	CS	QLT	Both	RE	A/B/C/D	P	Y	P	Y	Y	N
[87]	I	N/A	No	CS	QLT	Safety	RA/RE	A	P	Y	P	Y	N	N

Table 1. Attributes and Characteristics

Table 1 depicts both the attributes and the characteristics of all methods reviewed in the surveys of Section 2.1 and of those reviewed in Section 2.2.2.

4. Discussion

The main findings from the analysis of the literature reviewed in the previous section are the following:

- A total of sixty eight methods have been reviewed. These span a time period of approximately 20 years, with most having been proposed after 2013, and with a steady increase in the past 5 years. This is an indication of the timeliness of the subject, which can be attributed to the increased proliferation of cyber physical systems and the integration of Information Technology with Operational Technology.
- The number of integrated methods (37) is slightly larger than that of unified ones (31). According to Reference [62], approaches that attempt to unify safety and security analysis techniques reduce

the developer's understanding of the system being analyzed and prevent a thorough analysis of either property; this leads to an incomplete analysis with subsequent safety and security risks going unobserved. On the other hand, integrated methods extract more rigorous results and facilitate the identification of potential conflicts.

- Model-based methods prevail (52 out of 68). Of these, 18 methods employ formal models, 23 methods employ graphical models, and 11 methods employ both formal and graphical models. Model-based approaches are more practical for modeling a system's components and functionalities for existing and operational systems, by virtue of their qualitative and quantitative capabilities [4]. They are generally able to scale up to complex systems and represent different aspects related to safety and security with different viewpoints and levels of detail. On the other hand, such approaches require the analyst to have a thorough knowledge of the system; engaging all stakeholders in the process my facilitate the fulfillment of this requirement.
- Less than half of the reviewed methods (20) are informed by safety and security standards. Cyberphysical systems often operate in domains and environments highly regulated by safety and security standards. Therefore, they must be engineered to conform to these standards. It follows that safety-security co-engineering methods needs to be informed by standards. This need is more often than not satisfied if the method has been designed for use in a specific application domain. Including a validation phase in the workflow of the method, in which conformance to relevant standards is performed, is a viable alternative that may lead to the development of generic methods informed by standards. A related issue, discussed in the next section, is the need for integrated safety-security standards in several application domains.
- Most (45) of the reviewed methods have been used to analyze general CPS architectures and industrial control systems in various application domains, with the transportation domain prevailing. However, the applicability of the generic methods to different application domains is usually not demonstrated. Developing a method applicable to a broad spectrum of domains and at the same time ensuring compliance with relevant standards appears as a very challenging task.
- The vast majority of the reviewed methods (66) follow -at least partially- a qualitative approach; only two methods are fully quantitative. This is not surprising, because even though quantitative approaches prevail for safety engineering, the opposite is true for security engineering, where quantitative approaches are very rarely used, as they require the existence of a formal model describing the system under study. Attempting to analyze security, particularly security risk, has been shown quantitatively to be either infeasible or inadvisable in most real-world situations. Hence, a reasonable compromise is to opt for a combination of quantitative and qualitative approaches for safety-security co-engineering.
- The number of methods whose goal is to ensure both safety and security (32) is slightly larger than the number of those aiming to ensure safety (30), whilst only six methods have as their primary goal to ensure security. The appropriateness of each of these approaches depends largely on the system's safety/security criticality. When the system under study is safety critical, a method whose goal is to ensure that security will not adversely influence safety is appropriate; the opposite is true when the system is security critical. But if the intention is to also have a secure system beyond the safety relevant security issues, and a safe system beyond the security relevant issues, then neither of these approaches is appropriate. In systems where both safety and security are equally important, an approach aiming to ensure both safety and security would be more appropriate.
- The number of reviewed methods that are applied to both the requirements elicitation and risk analysis phases of the system lifecycle (26) is almost equal to that of methods applied to the risk analysis phase (25), and only slightly exceeds that of methods applied to the requirements elicitation phase (17); only fifteen methods are frameworks, hence applicable to any phase of the lifecycle. This nearly uniform distribution reflects the emphasis given into co-engineering safety

and security as early as possible in the development lifecycle, while allowing for revisiting the results of the analysis when the system under study has been developed or is even operational.

- The application of most of the reviewed methods involves safety and security experts; only few methods require the engagement of developers, designers and systems users. It is important to note that stakeholders, particularly designers and users, may engage with the analysis in two distinct but complementary ways: they provide input to the analysis in the form of domain expert knowledge, and they are the targets of the process of communicating the results; both are equally important. Acknowledging the fact that complex issues such as those of safety and security cannot be effectively analysed by the corresponding experts, it follows that successful methods will seek to involve engaged stakeholders.
- Scalability issues have been discussed in the vast majority (61) of the papers proposing methods, in 24 of which these issues have only briefly been considered. Thirty seven methods are scalable, whereas 7 do not scale well. It should be noted that scalability refers to both the ability of the method to handle complex systems and to the level of abstraction at which the system under study is represented. The two are correlated, as high level abstraction allows for more complex systems to be analyzed. The challenge, therefore, is to develop methods that can strike an appropriate balance between those two aspects of scalability, so that the analysis results in an appropriate and practically useful level of detail.
- The majority (55) of the reviewed methods provide mechanisms to stimulate creativity among the analysts and other relevant stakeholders. As many methods rely, at least to some extent, on scenario development, creativity is an important characteristic. This is even more so when the application of the method calls for a multi-disciplinary, multi-stakeholder team.
- Less than half (28) of the reviewed methods include techniques to communicate their results to the relevant stakeholders. Another 28 methods only briefly address the issue, whilst 10 methods do not address it at all. As already implied above, this characteristic is intertwined with the involvement of stakeholders attribute.
- All methods are process-based; the structure and the steps of the process do vary, however. As pointed out in the sequel, developing a methodology to encompass different process structures is still a challenge.
- The majority (49) of the reviewed methods do not address the conflict resolution issue. Sixteen methods do address it, and a further 3 address it only partially. The implications of this central issue is elaborated upon in the sequel.
- The majority (41) of the reviewed methods are not supported by any software tool or toolkit. Only 20 methods are fully supported, and another 7 are partially supported by such tools. This has been a rather surprising finding, as the purpose of a safety-security co-engineering method is to be applied in real-world application scenarios. The complexity of such methods requires software support for their usage.

To give a bird's eye view of these findings, and also to facilitate cross-referencing, the above are summarized in Figures 2 and 3. Figure 2 depicts the taxonomy of Figure 1, with the number following each attribute indicates the number of methods having the corresponding attribute. Figure 3 provides the same information on characteristics.

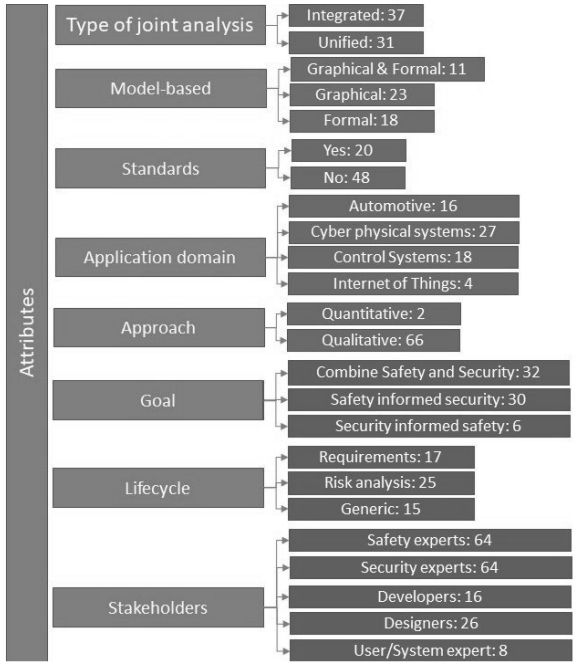


Figure 2. Attributes: Results

Figure 3. Characteristics: Results

Additionally, a number of issues that have been under-researched have been identified. These are as follows:

- *Resolving conflicting safety/security results:* The problem of conflicting results when studying safety and security jointly has been known for some time. There are two approaches to address this problem: either allow conflicting results to be derived and then resolve these conflicts, or avoid the occurrence of such conflicts by design. Unified safety and security co-engineering methods tend to generate less conflicts than integrated methods do. However, integrated methods tend to allow more comprehensive analyses of both domains. Therefore, integrated methods that would by design prevent the occurrence of conflicting results would address this issue effectively. This could perhaps be achieved if goal oriented integrated methods were developed. Further, the analysis is best performed in the early stages (requirements elicitation phase), as this makes the problem of conflict resolution much easier to solve, and leads to the development of safe-and-secure CPSs by design.
- *Standard methodology:* Despite the sizable extent of the literature on safety and security co-engineering methods, a generic, application-domain-independent *methodology*, instances of which would be existing methods and those to be developed in the future, is yet to be developed. An example of such a methodology in the security domain is the risk analysis methodology defined in the ISO 27005 standard.

- *Validation:* Not many of the reviewed papers include information on the validation/evaluation of the method they propose. More research is needed to evaluate the correctness, completeness, effectiveness, efficiency, scalability and so forth, of proposed methods, in a manner that will facilitate comparative assessments.
- *Safety and security standards:* Some standards addressing safety and security for industrial control systems exist. Examples of such standards are ISA99/IEC 6443, IEC 62645, IEC TR63609, ISO 26262 to name a few; cross-references with other standards (e.g., IEC 61508) also exist. However, the applicability of such standards to effectively address both safety and security, particularly in an industry 4.0 context, is still to be firmly established. Hence, a need for revisiting existing standards with an eye towards facilitating their use in industry, by means of reducing ambiguity, arises. Additionally, the adoption of standards specific for industry sectors, along the lines of the practice followed in the nuclear plant domain will guide the development of safe-and-secure by design industrial control systems.
- *Application domains:* As noted before, the transportation domain prevails among application domains addressed by the reviewed methods. Notwithstanding the fact that several methods have been claimed to have been designed to be applicable to any domain, their applicability has not been demonstrated. As several emerging application domains are both safety and security critical (e.g., autonomous vessels, drones), the development of methods addressing specifically systems in such domains remains an issue.
- *Dynamic character of CPS:* CPSs are dynamic by nature. Methods able to model and cope with this characteristic of CPSs are yet to be developed. Existing work on dynamic security and dynamic safety risk assessment can be leveraged to this end.
- *Model type:* Most of the safety analysis approaches are based on formal models. Security techniques on the other hand tend to focus on qualitative analysis. Therefore, an approach able to handle the complexity of CPS by leveraging both graphical models and systematic perspectives would allow the consolidation of advantages of both worlds.
- *Holistic approach:* The human factor in relation with CPSs is often overlooked. In fact, CPSs, particularly safety/security critical ones need to be considered and studied as socio-technical systems. This calls for a holistic approach towards safety and security co-engineering, that would encompass the whole ecosystem into which the CPS under study is expected to operate, and would involve all the relevant stakeholders in the process. To this end, future methods should enjoy previously mentioned attributes such as *scalability*, *communication*, and *model type*, in order to facilitate the analysis of CPSs when both technical and human aspects are considered. Particularly, such methods should be able to handle the complexity (*scalability*) derived from the human-machine interaction; communicate the results by providing reports and leveraging software tools (*communication*); and provide graphical models of the system under study (*model type*) to facilitate the analysis and the validation of the results.

5. Conclusions

We have revisited previous surveys on cybersecurity and safety co-engineering approaches and performed a systematic literature survey of such approaches. We defined a multi-attribute taxonomy for such approaches and we used this to analyze them. We thus provided a comprehensive discussion on the recent advances in cybersecurity and safety co-engineering. The joint study of safety and security has been a goal of researchers in both fields for more than thirty years. Despite the longevity of the problem and the substantial volume of research results on safety and security co-engineering that has been generated in the past few years, several important issues remain open. Through our review, we identified and discussed such issues and the research challenges that they imply. In the future, among the many possible research challenges in the field, we plan to focus on developing a holistic, integrated, graphical model based, safety and security requirements elicitation co-engineering approach, applicable to the autonomous vessel domain.

Author Contributions: Conceptualization, G.K. and S.K.; methodology, G.K.; investigation, G.K.; writing–original draft preparation, G.K.; writing–review and editing, S.K. and G.K.; supervision, S.K. and V.G.; project administration, S.K. and V.G.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pièrre-Cambacédès, L.; Bouissou, M. Modeling safety and security interdependencies with BDMP (Boolean logic Driven Markov Processes). In Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 2852–2861.
2. Paul, S.; Rioux, L. *Over 20 Years of Research into Cybersecurity and Safety Engineering: a Short Bibliography*; WIT Press: Ashurst Lodge, UK, 2015, pp. 335–349. doi: 10.2495/SAFE150291.
3. Pièrre-Cambacédès, L.; Bouissou, M. Cross-fertilization between safety and security engineering. *Reliab. Eng. Syst. Safe.* **2013**, *110*, 110–126.
4. Kriaa, S.; Pietre-Cambacedes, L.; Bouissou, M.; Halgand, Y. A survey of approaches combining safety and security for industrial control systems. *Reliab. Eng. Syst. Safe.* **2015**, *139*, 156–178.
5. Chockalingam, S.; Hadžiosmanović, D.; Pieters, W.; Teixeira, A.; van Gelder, P. Integrated safety and security risk assessment methods: a survey of key characteristics and applications. In Proceedings of the International Conference on Critical Information Infrastructures Security, Paris, France, 10–12 October 2016, pp. 50–62.
6. Abulamddi, M.F. A Survey on techniques requirements for integrating safety and security engineering for cyber-physical systems. *J. Comput. Commun.* **2017**, *5*, 94–100.
7. Lisova, E.; Sljivo, I.; Causevic, A. Safety and Security Co-Analyses: A Systematic Literature Review. *IEEE Syst. J.* **2018**, *13*, 2189–2200. doi: 10.1109/JSYST.2018.2881017.
8. Lyu, X.; Ding, Y.; Yang, S.H. Safety and security risk assessment in cyber-physical systems. *IET Cyber-Physical Syst. Theory Appl.* **2019**, *4*, 221–232. doi: 10.1049/iet-cps.2018.5068.
9. Hart, C. *Doing a literature review: Releasing the research imagination*; SAGE Publications Ltd: Southend Oaks, CA, USA, 2018.
10. Cui, J.; Sabaliauskaite, G. US 2 : An Unified Safety and Security Analysis Method for Autonomous Vehicles. In Proceedings of the Future of Information and Communication Conference, Singapore, Singapore, 5–6 April 2018, pp. 600–611.
11. Sabaliauskaite, G.; Liew, L.S.; Cui, J. Integrating autonomous vehicle safety and security analysis using stpa method and the six-step model. *Int. J. Adv. Secur.* **2018**, *11*, 160–169.
12. Sabaliauskaite, G.; Adepu, S.; Mathur, A. A six-step model for safety and security analysis of cyber-physical systems. International Conference on Critical Information Infrastructures Security, Paris, France, 10–12 October 2016, pp. 189–200.
13. Leveson, N.G.; Thomas, J.P. *STPA handbook*; USA**2018**, Available online:https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf.
14. International Organization for Standardization (ISO). *Road vehicles — Functional safety*; Technical report ISO 26262-1:2018. ISO, 2018.
15. SAE, J. 3061: Cybersecurity guidebook for cyber-physical vehicle systems, 2016. Available online: https://www.sae.org/standards/content/j3061/ (accessed on 18-10-2019)
16. Sabaliauskaite, G.; Mathur, A.P. Aligning cyber-physical system safety and security, 2015, Available online:http://www.2014.csdm-asia.net/IMG/pdf/Aligning_Cyber-Physical_System_Safety_and_Security-2.pdf (accessed on 20-11-2019)
17. Cui, J.; Sabaliauskaite, G. On the alignment of safety and security for autonomous vehicles. In Proceedings of the CYBER 2017 : The Second International Conference on Cyber-Technologies and Cyber-Systems, IARIA CYBER, Barcelona, Spain, 12–16 November 2017.
18. International Society of Automation - ISA. Technical report, ANSI/ISA 84.00.01-2004, Application of Safety Instrumented Systems for the Process Industries. publisher: ISA, 2004.
19. International Society of Automation - ISA. Technical report, ANSI/ISA-99-00-01-2007. Security for Industrial Automation and Control Systems Part 1: Terminology, Concepts, and Models. publisher: ISA, 2007.

20. Asplund, F.; McDermid, J.; Oates, R.; Roberts, J. Rapid Integration of CPS Security and Safety. *IEEE Embed. Syst. Lett.* **2018**, *11*, 111–114. doi:10.1109/LES.2018.2879631.
21. Guzman, N.H.C.; Kufoalor, D.K.M.; Kozin, I.; Lundteigen, M.A. 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, Hannover, Germany, 22–26 September 2019, pp. 4099–4106.
22. Sabaliauskaite, G.; Liew, L.S.; Zhou, F. AVES–Automated Vehicle Safety and Security Analysis Framework. In Proceedings of the CSCS '19: ACM Computer Science in Cars Symposium ,Kaiserslautern, Germany, 8 October 2019, pp. 1–8. doi: 10.1145/3359999.3360494.
23. Guzman, N.H.C.; Wied, M.; Kozine, I.; Lundteigen, M.A. Conceptualizing the key features of cyber-physical systems in a multi-layered representation for safety and security analysis. *Syst. Eng.* **2019**, *23*, 189–210.
24. Carreras Guzman, N.H.; Mezovari, A.G. Design of IoT-based Cyber–Physical Systems: A Driverless Bulldozer Prototype. *Information* **2019**, *10*, 343.
25. Hayakawa, T.; Sasaki, R.; Hayashi, H.; Takahashi, Y.; Kaneko, T.; Okubo, T. Proposal and Application of Security/Safety Evaluation Method for Medical Device System that Includes IoT. In Proceedings of the 2018 VII International Conference on Network, Communication and Computing, Taipei, Taiwan, 14–16 December 2018, pp. 157–164.
26. Monteuuis, J.P.; Boudguiga, A.; Zhang, J.; Labiod, H.; Servel, A.; Urien, P. Sara: Security automotive risk analysis method. In Proceedings of the 4th ACM Workshop on Cyber-Physical System Incheon Republic of Korea 2018, pp. 3–14.
27. Raspotnig, C.; Opdahl, A. Comparing risk identification techniques for safety and security requirements. *J. Syst. Softw.* **2013**, *86*, 1124–1151.
28. Raspotnig, C.; Karpati, P.; Katta, V. A combined process for elicitation and analysis of safety and security requirements. *Enterprise, Business-process and Information Systems Modeling*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 347–361.
29. Reichenbach, F.; Endresen, J.; Chowdhury, M.M.; Rossebø, J. A pragmatic approach on combined safety and security risk analysis. In Proceedings the 2012 IEEE 23rd International Symposium on Software Reliability Engineering Workshops, Dallas, TX, USA, 27–30 November 2012; pp. 239–244. doi: 10.1109/ISSREW.2012.98.
30. Silva, N.; Lopes, R. Practical Experiences with real-world systems: Security in the World of Reliable and Safe Systems. 2013 43rd Annual IEEE/IFIP Conference on Dependable Systems and Networks Workshop (DSN-W), Budapest, Hungary, 24–27 June 2013, pp. 1–5.
31. Young, W.; Leveson, N. Systems thinking for safety and security. Proceedings of the 29th Annual Computer Security Applications Conference, New Orleans Louisiana USA, December 2013, pp. 1–8.
32. Chen, Y.R.; Chen, S.J.; Hsiung, P.A.; Chou, I.H. Unified security and safety risk assessment-a case study on nuclear power plant. In Proceedings of the 2014 International Conference on Trustworthy Systems and Their Applications, Taichung, Taiwan, 9–10 June 2014, pp. 22–28.
33. Ito, M. Finding threats with hazards in the concept phase of product development. European Conference on Software Process Improvement, Luxembourg, Luxembourg, 25–27 June 2014, pp. 277–284.
34. Kriaa, S.; Bouissou, M.; Colin, F.; Halgand, Y.; Pietre-Cambacedes, L. Safety and security interactions modeling using the BDMP formalism: case study of a pipeline. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, SAFECOMP 2014, Florence, Italy, September 2014, pp. 326–341.
35. Schmittner, C.; Gruber, T.; Puschner, P.; Schoitsch, E. Security application of failure mode and effect analysis (FMEA). In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Florence, Italy, September 2014, pp. 310–325.
36. Apvrille, L.; Roudier, Y. Designing safe and secure embedded and cyber-physical systems with SysML-Sec. In Proceedings of the International Conference on Model-Driven Engineering and Software Development, Angers, France, 9–11 February 2015, pp. 293–308.
37. Gu, T.; Lu, M.; Li, L. Extracting interdependent requirements and resolving conflicted requirements of safety and security for industrial control systems. 2015 First International Conference on Reliability Systems Engineering (ICRSE), Beijing China, October 2015, pp. 1–8.
38. Kriaa, S.; Bouissou, M.; Laarouchi, Y. A model based approach for SCADA safety and security joint modelling: S-Cube. In Proceedings of the 10th IET System Safety and Cyber-Security Conference 2015, Bristol, UK, 21–22 October 2015.

39. Macher, G.; Höller, A.; Sporer, H.; Armengaud, E.; Kreiner, C. A combined safety-hazards and security-threat analysis method for automotive systems. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Delft, The Netherlands, 22–25 September 2014, pp. 237–250.
40. Popov, P.T. Stochastic Modeling of Safety and Security of the e-Motor, an ASIL-D Device. In *Computer Safety, Reliability, and Security*; Koornneef, F.; van Gulijk, C., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 385–399.
41. Steiner, M.; Liggesmeyer, P. Qualitative and quantitative analysis of CFTs taking security causes into account. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Delft, The Netherlands, 22–25 September 2014, pp. 109–120.
42. Wei, J.; Matsubara, Y.; Takada, H. HAZOP-Based Security Analysis for Embedded Systems: Case Study of Open Source Immobilizer Protocol Stack. In *Recent Advances in Systems Safety and Security*; Springer International Publishing: Cham, Switzerland, 2016; pp. 79–96.
43. Islam, M.M.; Lautenbach, A.; Sandberg, C.; Olovsson, T. A risk assessment framework for automotive embedded systems. In Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security, Xi'an China, 30 May 2016, pp. 3–14.
44. Nicklas, J.P.; Mamrot, M.; Winzer, P.; Lichte, D.; Marchlewitz, S.; Wolf, K.D. Use case based approach for an integrated consideration of safety and security aspects for smart home applications. In Proceedings of the 2016 11th System of Systems Engineering Conference (SoSE). Kongsberg, Norway, 12–16 June 2016.
45. Ponsard, C.; Dallons, G.; Massonet, P. Goal-oriented co-engineering of security and safety requirements in cyber-physical systems. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Trondheim, Norway, 20–23 Septembe 2016, pp. 334–345.
46. Schmittner, C.; Ma, Z.; Puschner, P. Limitation and improvement of STPA-Sec for safety and security co-analysis. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Trondheim, Norway, 20–23 Septembe 2016, pp. 195–209.
47. Troubitsyna, E. An integrated approach to deriving safety and security requirements from safety cases. In Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), Atlanta, GA, USA, 10–14 June 2016; pp. 614–615.
48. Dürrwang, J.; Beckers, K.; Kriesten, R. A lightweight threat analysis approach intertwining safety and security for the automotive domain. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Trento, Italy, 12–15 September 2017; pp. 305–319.
49. Friedberg, I.; McLaughlin, K.; Smith, P.; Laverty, D.; Sezer, S. STPA-SafeSec: Safety and security analysis for cyber-physical systems. *J. Inf. Secur. Appl.* **2017**, *34*, 183–196.
50. Howard, G.; Butler, M.; Colley, J.; Sassone, V. Formal analysis of safety and security requirements of critical systems supported by an extended STPA methodology. 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Paris, France, 26–28 April 2017, pp. 174–180.
51. Kumar, R.; Stoelinga, M. Quantitative security and safety analysis with attack-fault trees. 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE), Singapore, Singapore, 12–14 January 2017; pp. 25–32.
52. Pereira, D.; Hirata, C.; Pagliares, R.; Nadjm-Tehrani, S. Towards combined safety and security constraints analysis. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Trento, Italy, 12–15 September 2017, pp. 70–80.
53. Plósz, S.; Schmittner, C.; Varga, P. Combining safety and security analysis for industrial collaborative automation systems. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Trento, Italy, 12–15 September 2017; pp. 187–198.
54. Procter, S.; Vasserman, E.Y.; Hatcliff, J. SAFE and secure: Deeply integrating security in a new hazard analysis. In Proceedings of the 12th International Conference on Availability, Reliability and Security, Reggio Calabria, Italy, 29 August –1 September 2017, p. 66. doi: 10.1145/3098954.3105823.
55. Sabaliauskaite, G.; Adepu, S. Integrating six-step model with information flow diagrams for comprehensive analysis of cyber-physical system safety and security. 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE), Singapore, Singapore, 12–14 January 2017, pp. 41–48.
56. Temple, W.G.; Wu, Y.; Chen, B.; Kalbarczyk, Z. Systems-theoretic likelihood and severity analysis for safety and security co-engineering. In Proceedings of the International Conference on Reliability, Safety and Security of Railway Systems, Italy, 12–15 September 2017; pp. 51–67.

57. Vistbakka, I.; Troubitsyna, E.; Kuismin, T.; Latvala, T. Co-engineering safety and security in industrial control systems: a formal outlook. In Proceedings of the International Workshop on Software Engineering for Resilient Systems, Geneva, Switzerland, 4–5 September 2017, pp. 96–114.
58. Stoneburner, G. Toward a unified security-safety model. *Computer* **2006**, *39*, 96–97.
59. Aven, T. A unified framework for risk and vulnerability analysis covering both safety and security. *Reliab. Eng. Syst. Safe.* **2007**, *92*, 745–754.
60. Derock, A.; Hebrard, P.; Vallée, F. Convergence of the latest standards addressing safety and security for information technology, 2010. Available online: https://hal.archives-ouvertes.fr/hal-02267717/ (accessed on 15-10-2019)
61. Woskowski, C. A pragmatic approach towards safe and secure medical device integration. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Florence, Italy, September 2014, pp. 342–353.
62. Eames, D.P.; Moffett, J. The integration of safety and security requirements. In Proceedings of the International Conference on Computer Safety, Reliability, and Security, Toulouse, France, 27–29 September 1999, pp. 468–480.
63. Kornecki, A.J.; Zalewski, J. Safety and security in industrial control. In Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research, Oak Ridge, TN, USA 21–23 April 2010, p. 77. doi: 10.1145/1852666.1852754.
64. Novak, T.; Gerstinger, A. Safety-and security-critical services in building automation and control systems. *IEEE Trans. Ind. Electron.* **2009**, *57*, 3614–3621.
65. Subramanian, N.; Zalewski, J. Assessment of safety and security of system architectures for cyberphysical systems. In Proceedings the 2013 IEEE International Systems Conference (SysCon), Orlando, FL, USA, 15–18 April 2013, pp. 634–641. doi: 10.1109/SysCon.2013.6549949.
66. Fovino, I.N.; Masera, M.; De Cian, A. Integrating cyber attacks within fault trees. *Reliab. Eng. Syst. Safe.* **2009**, *94*, 1394–1402.
67. Bezzateev, S.; Voloshina, N.; Sankin, P. Joint safety and security analysis for complex systems. In Proceedings the 2013 13th Conference of Open Innovations Association (FRUCT), Petrozavodsk, Russia, 22–26 April 2013, pp. 3–13.
68. Kornecki, A.; Liu, M. Fault tree analysis for safety/security verification in aviation software. *Electronics* **2013**, *2*, 41–56.
69. Steiner, M.; Liggesmeyer, P. Combination of safety and security analysis-finding security problems that threaten the safety of a system, 2013. Available online: https://hal.archives-ouvertes.fr/hal-00848604 (accessed on 04-11-2019)
70. Pièrre-Cambacédès, L.; Deflesselle, Y.; Bouissou, M. Security modeling with BDMP: from theory to implementation. 2011 Conference on Network and Information Systems Security, La Rochelle, France, 18–21 May 2011; doi: 10.1109/SAR-SSI.2011.5931382.
71. Kornecki, A.J.; Subramanian, N.; Zalewski, J. Studying interrelationships of safety and security for software assurance in cyber-physical systems: Approach based on bayesian belief networks. 2013 Federated Conference on Computer Science and Information Systems, Krakow, Poland, 8–11 September 2013, pp. 1393–1399.
72. Sindre, G. A Look at Misuse Cases for Safety Concerns. *Situational Method Engineering: Fundamentals and Experiences*; Ralyté, J.; Brinkkemper, S.; Henderson-Sellers, B., Eds.; Springer US: Boston, MA, USA, 2007; pp. 252–266.
73. Jürjens, J. *Developing safety-and security-critical systems with UML*. DARP workshop: Loughborough, UK, 2003.
74. Apvrille, L.; Roudier, Y. Towards the model-driven engineering of secure yet safe embedded systems.*arXiv* **2014** *arXiv preprint arXiv:1404.1985*. Available online: https://arxiv.org/abs/1404.1985 (accessed on 08-11-2019)
75. Roth, M.; Liggesmeyer, P. Modeling and analysis of safety-critical cyber physical systems using state/event fault trees. Available online: https://hal.archives-ouvertes.fr/SAFECOMP2013-DECS/hal-00848640 (accessed on 18-10-2019)

76. Brunel, J.; Chemouil, D.; Rioux, L.; Bakkali, M.; Vallée, F. A viewpoint-based approach for formal safety & security assessment of system architectures. Available online: https://hal.archives-ouvertes.fr/hal-01070960 (accessed on 22-10-2019)
77. Zafar, S.; Dromey, R.G. Integrating safety and security requirements into design of an embedded system. 12th Asia-Pacific Software Engineering Conference (APSEC'05), Taipei, Taiwan, 15–17 December 2005; pp. 8–pp.
78. Pieters, W.; Lukszo, Z.; Hadziosmanovic, D.; van den Berg, J. Reconciling Malicious and Accidental Risk in Cyber Security. *J. Internet Serv. Inf. Secur.* **2014**, *4*, 4–26.
79. Sun, M.; Mohan, S.; Sha, L.; Gunter, C. Addressing safety and security contradictions in cyber-physical systems. In Proceedings of the 1st Workshop on Future Directions in Cyber-Physical Systems Security (CPSSW'09), Newark, NJ,USA, 22–24, July 2009. doi:10.1049/iet-cps.2018.5068.
80. Simpson, A.; Woodcock, J.; Davies, J. Safety through security. In Proceedings Ninth International Workshop on Software Specification and Design, Washington, DC, USA, 1998, pp. 18–24.
81. Delange, J.; Pautet, L.; Feiler, P. Validating safety and security requirements for partitioned architectures. In Proceedings of the International Conference on Reliable Software Technologies, Brest, France,8–12 June 2009, pp. 30–43.
82. Young, W.; Leveson, N.G. An integrated approach to safety and security based on systems theory. *Commun. ACM* **2014**, *57*, 31–35.
83. 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 Proceedings of the 1st ACM Workshop on Cyber-Physical System Security, Singapore, Singapore, 14 April 2015, pp. 69–80.
84. Schmittner, C.; Ma, Z.; Smith, P. FMVEA for safety and security analysis of intelligent and cooperative vehicles. In Proceeddings of the International Conference on Computer Safety, Reliability, and Security, Florence, Italy, September 2014, pp. 282–288.
85. Chung, L.; Nixon, B.A.; Yu, E.; Mylopoulos, J. *Non-functional Requirements in Software Engineering*; Springer US: Cham, Switzerland; 2012.
86. Subramanian, N.; Zalewski, J. Quantitative assessment of safety and security of system architectures for cyberphysical systems using the NFR approach. *IEEE Syst. J.* **2014**, *10*, 397–409.
87. Winther, R.; Johnsen, O.A.; Gran, B.A. Security assessments of safety critical systems using HAZOPs. International Conference on Computer Safety, Reliability, and Security, Budapest, Hungary, 26–28 September 2001, pp. 14–24.

9 Article V: SafeSec Tropos: Joint security and safety requirements elicitation [5]

SafeSec Tropos: Joint security and safety requirements elicitation

Computer Standards Interfaces, Volume 70, 2020, 103429, ISSN 0920-5489,

https://doi.org/10.1016/j.csi.2020.103429.

Georgios Kavallieratos [a], Sokratis Katsikas [a,b] and Vasileios Gkioulos [a]

[a] Norwegian University of Science and Technology, Department of Information Security and Communications Technology, Gjøvik, Norway
[b] Open University of Cyprus, School of Pure and Applied Sciences, Latsia, Nicosia, Cyprus

ARTICLE INFO

Keywords:
security
safety
cyber physical systems
requirements elicitation
maritime ecosystem

ABSTRACT

The growing convergence of information technology with operational technology and the accordant proliferation of interconnected cyber-physical systems (CPSs) has given rise to several security and safety challenges. One of these refers to systematically identifying coherent, consistent, and non-conflicting security and safety requirements. This paper proposes an integrated method for safety and security requirements engineering for CPSs at the design stage of the system lifecycle. The method identifies security and safety objectives, it systematically elicits a comprehensive list of requirements, and it links these requirements to objectives, thus facilitating the process of resolving conflicts. To provide insight into the operations of the method, we demonstrate its use to the most vulnerable CPSs on board the Cyber-Enabled Ship (C-ES). By utilizing the proposed method, the safety and security objectives of these systems were defined, and their safety and security requirements were identified.

1. Introduction

Due to the close intertwining of the cyber and physical components, both safety and security are essential for the reliable operation of cyber-physical systems (CPSs). Safety aims to protect systems from unintentional actions while security implies protection from both intentional and unintentional threats. The associations between safety and security have extensively been analyzed in the literature [25, 51, 55].

Requirements engineering (RE) is a vital element of the CPS development process. As such, it is incorporated in both the safety [19] and the security [22] lifecycles and it is described in both safety and security standards. Several standards on the security and safety of cyber physical systems are discussed in [2, 57]. These include the ISO 27k family; NEC's CIP family of standards; and the ISA IEC IEC-62443 series. Also relevant are standards on software security requirements (such as e.g. ECSS-Q-ST-80 C, IEEE 830-1998, ISO/IEC 25010, ISO/IEC 27034-1, and ISO/IEC 27034-3). From the safety point of view, various standards exist for safety in the maritime domain. Such standards have been developed by ISO/TC 8/SC 1 and the International Maritime Organization (IMO). Additionally, various standards have been surveyed in [37] regarding the functional safety and security of industrial control systems. The incorporation of safety and security aspects is discussed in IEC 62859 for nuclear power plants. However, standards regarding the co-analysis of safety and security or even security alone in the maritime domain have not yet been developed. RE is incorporated in both the safety [19] and the security [22] lifecycles. As a weak combination of safety and security requirements may result in poor system design and development and possibly to damages to the CPS ecosystem, jointly analyzing and eliciting requirements for security and safety is necessitated. This is particularly so in cases where increasingly complex and interconnected CPSs are utilized, such as the Cyber-Enabled ship (C-ES) case. The C-ES is a variant of the autonomous or remotely controlled vessel. Its operational and functional activities are described considering Autonomy Levels (AL) 1-3, of the International Maritime Organization (IMO) classification [21]. As discussed in detail in Section 2, various security and safety requirements engineering methodologies have appeared in the literature. However, several impediments are associated with the co-analysis of security and safety in complex systems [55]. Most of the existing studies fall short of eliciting requirements that ensure both the satisfaction

georgios.kavallieratos@ntnu.no (G.K.); sokratis.katsikas@ntnu.no (S.K.); vasileios.gkioulos@ntnu.no (V.G.)
ORCID(s): 0000-0003-1278-1943 (G.K.); 0000-0003-2966-9683 (S.K.); 0000-0001-7304-3835 (V.G.)

of safety constraints and the protection of information security attributes and are therefore not readily applicable to the security and safety co-analysis of CPSs within the C-ES ecosystem.

This work proposes an integrated method for the joint elicitation of security and safety requirements early during system design. The introduced method examines the safety and the security of the targeted system by analyzing the corresponding objectives. This process enables the identification of requirements early in the system's design phase, and facilitates the resolution of potential conflicts between safety and security requirements. In particular, the evaluation and selection of the identified objectives, and the consideration of the relevant standards [19, 22], legislation, and stakeholders, facilitate the requirements elicitation process, where the safety and the security objectives are translated to corresponding requirements. Moreover, the elicitation of the safety and security requirements is not performed in isolation; this is a common weakness of other integrated methodologies [37]. The method analyzes both safety and security objectives evenly, by integrating two well established systematic approaches, namely the Secure Tropos [42] and the Systems Theoretic Process Approach (STPA) [33]. The analysis follows a top-down approach and allows the extraction of results without the need to consider the detailed design of the system under analysis. Thus, the proposed method enables the analysis of systems whose detailed specifications are not yet available. We then apply the proposed method to the use case of the C-ES to identify safety and security requirements for the most vulnerable onboard CPS, namely the Automatic Identification System (AIS), the Electronic Chart Display System (ECDIS), and the Global Maritime Distress and Safety System (GMDSS). The contribution of this work is threefold:

- a novel method for the joint elicitation of security and safety requirements of CPSs has been developed based on the Secure Tropos and STPA methods from the security and the safety domain respectively;
- a set of security and safety objectives for the C-ES ecosystem has been defined. The security objectives are based on the Parkerian Hexad with the addition of Non-Repudiation. To the best of our knowledge, this work is a first attempt to define safety objectives for the CPSs of the C-ES's ecosystem;
- the security and safety requirements of the most vulnerable navigational CPSs onboard a C-ES (AIS, ECDIS, GMDSS) have been identified.

The remainder of this paper is structured as follows: Section 2 reviews the related work. Section 3 provides an overview of the Secure Tropos and the STPA methods. Section 4 discusses the limitations of the existing approaches and describes the proposed SafeSec Tropos method. Section 5 presents the application of the proposed method to the C-ES case. Section 6 discusses the results as well as limitations of the proposed method. Finally, 7 summarizes our conclusions and discusses challenges to be addressed in future research.

2. Related Work

The joint analysis of safety and security has received considerable attention. As a result, several relevant works exist in the literature, as well as reviews and surveys. A systematic literature review of safety and security co-analysis methods appeared in [35]; risk assessment approaches for security and safety of CPSs are surveyed in [38]; approaches combining security and safety for industrial control systems are surveyed in [31]; and safety and security co-engineering methods are surveyed in [47]. C. Raspotnig et al. in [53] surveyed risk analysis methods for safety and security and compared the surveyed techniques considering twelve criteria. The survey concluded that there is a need for a tighter integration of the requirements elicitation activities with safety and security aspects.

Two approaches regarding the co-analysis of safety and security are identified: (1) Integrated approach, and (2) Unified approach. An integrated approach analyzes safety and security separately and then integrates the results, while a unified approach analyzes safety and security jointly [35]. The former reduces the insight of the analysis leading to incomplete results, while the latter provides more rigorous results, with better understanding of potential conflicts between safety and security [12]. The existing methods have different characteristics depending on their approach towards analyzing the security and safety of the targeted system (unified/integrated); the phase of system lifecycle when the method can be applied (Development/Operational); the approach towards identifying the requirements (Qualitative/Quantitative); and the way safety and security properties influence each other (safety informed security/ security informed safety/ combined safety and security) [35, 31].

Despite the diversity, most of the existing works on the joint analysis of safety and security tackle security only as a peripheral or constituent of safety, largely neglecting the necessity for ensuring the fulfilment of its distinct objectives. Further, a recurring problem with existing work on joint security and safety analysis is that it more often than not results in identifying conflicting requirements. A framework to detect conflicts between safety and security requirements early in the development phase was proposed in [59]. The mechanism relies on negotiating changes in the requirements

among safety and security engineers. A conflict resolution policy within the context of an approach based on the IEC 15408 and IEC 61508 standards was proposed in [45]. However, the proposed approach requires a formal description of the system under study, thus its applicability in practice is questionable. An approach to combine security and safety constraints by leveraging the NIST SP 800-30 standard and the Systems Theoretic Process Analysis (STPA) method -discussed in the next section- was proposed in [50], where potential conflicts among the requirements are resolved by either redefining the system or refining the requirements. However, to the best of our knowledge, a method that would jointly analyze safety and security, and allow the resolution of possible conflicts by means of prioritizing the objectives that generated each conflicting requirement has not been proposed.

The Cyber Risk Assesment Framework (CRAF) was proposed in [5] and was applied analyze the security and safety of a vessel. An STPA-based approach to enhance the co-analysis of security and safety and its application to a semi-autonomous vessel was presented in [16]. A method to combine security and safety during the risk analysis process of the collision avoidance function of an autonomous surface vessel was proposed in [17]. Safety-related cyber-attacks against the navigation and propulsion systems of an inland autonomous vessel were identified in [8]. Three safety and security co-analysis approaches using an autonomous boat as a case study were compared in [60]. Safety and security issues for the crewless merchant vessel, developed within the EU project MUNIN, are examined in [29, 30]. However, a systematic analysis of safety and security requirements of the C-ES ecosystem and its constituent CPSs has not been undertaken.

3. Background

Secure Tropos is a model-based method for security requirements engineering [42]. It facilitates the analysis of the system's environment, along with complex and distributed computerized systems, by using a graphical language. It encompasses four models, namely: the security reference model; the security constraint model; the security entities model; and the secure capability model. The method follows four stages: (1) The early requirements elicitation, where the actors, goals, assets, and resources of the whole ecosystem are identified. The outcome of this phase is an actor diagram and a number of goal diagrams. (2) The late requirements elicitation, where the actor diagram of the previous stage is extended with the introduction of the system under study as an actor that has a number of dependencies with the rest of the actors. (3) The architectural design where a global architecture of the system is defined. (4) The detailed design. These stages facilitate the security analysis of the targeted system by identifying the relevant stakeholders, system goals, processes, and activities. The use of the method is supported by a software tool (SecTro Tool) [48]. Secure Tropos has been applied in various domains. Although Secure Tropos is a well accepted method for security requirements elicitation, it does not provide for considering safety-related objectives; therefore it cannot support the elicitation of safety constraints and requirements. Additionally, unsafe control actions and safety constraints cannot be identified, since the methodology supports only the security analysis of the targeted system.

STPA [34] is a systemic approach for safety analysis, focusing on the control actions of each system. STPA is based on system theory and facilitates the analysis of the targeted ecosystem by considering system and software interdependencies. The goal of STPA is to prevent losses. The method identifies potential causes of accidents by considering safety as a system control (constraint) problem. P. Asare et al. in [4] discussed the fitness of the STPA for analyzing the safety of CPSs, since the method analyzes components which are cyber and physical with social boundaries. The STPA analysis starts with the identification of accident and loss events, followed by the definition of hazardous system states that are responsible for possible accidents. These hazards are refined as system safety constraints, aiming to prevent accidents from occurring. STPA is carried out in four steps: (1) Define the purpose of the analysis; (2) Model the control structure; (3) Identify Unsafe Control Actions; and (4) Identify loss scenarios. The identification of the safety constraints can be achieved by following the four STPA principles: (i) A control action required for safety is not provided or not followed; (ii) An unsafe control action is provided; (iii) A potential safe control action is provided too early or too late; and (iv) A control action required for safety is stopped too soon or applied for too long. These principles are depicted in STPA tables, where the safety constraints along with the unsafe control actions and their consequences are described. STPA is a hazard analysis technique based on system theory; thus, certain cyber security threats, such as e.g. threats against confidentiality or repudiation, are not analyzed, since such threats cannot influence the system's safety. Although various extensions of STPA aiming to address security aspects have been proposed [14, 64], they come with some limitations [56, 12]. In particular, STPA-Sec is not able to capture and analyze information disclosure issues, hence it cannot be used to study confidentiality aspects of the targeted system. Even though a new approach, called STPA-SafeSec, overcomes some of the limitations of STPA-Sec,

it is a unified safety and security analysis method; as such, it can lead to incomplete analysis [37].

Various reviews and surveys for security requirements engineering exist in the literature [44, 39, 43]. Secure Tropos is suggested for security requirements elicitation in [40, 43]. Further, Secure Tropos has been applied in different critical infrastructure domains to analyze cybersecurity aspects [52, 41]. As regards safety, various safety analysis techniques have been surveyed in [9, 53]. The survey concluded that the pros of the STPA are the wider perspective it provides on the system hazards; its ability to capture the control structure; and its coverage of conflicting actions in CPSs. An important advantage of STPA as compared to other safety analysis methods is that it considers the interactions among the system's components and it identifies safety constraints for such components [58]. This enables the analysis of more abstract systems whose technical and operational details have not been defined yet. Further, system hazards are identified in a more comprehensive way, based on the system's control structure [60]. Additionally, both Secure Tropos and STPA are top down approaches and facilitate the systemic analysis of the targeted system, early in the development phase. Accorddingly, Secure Tropos and STPA are chosen as the appropriate methods to combine in order to jointly study safety and security issues for CPSs.

This work proposes a method that addresses some of the identified limitations of existing alternatives:

- **Objectives-driven method:** To the best of our knowledge, none of the existing approaches analyzes a system and elicits requirements based on safety and security objectives. Doing so facilitates the *prioritization* of requirements, *communicating* them to relevant stakeholders, and *resolving conflicts* between safety and security requirements.
- **System models:** Security system models and safety system models can differ greatly, leading developers to take different views of the system, despite the fact that the underlying system is the same [12]. SafeSec Tropos allows the same model to be used for both safety and security analysis.
- **Documentation:** The documentation of the analyses can differ greatly between safety and security, thus making it difficult to find and compare requirements [12]. SafeSec Tropos provides similar documentation structures for both safety and security requirements, thus allowing easier identification of conflicts.
- **Conflict resolution:** The goal-oriented nature of SafeSec Tropos facilitates the resolution of potential conflicts between the safety and security requirements, as each requirement can be traced back to the objectives and goals that generated it.
- **Representation of complex systems:** SafeSec Tropos combines the graphical concepts of the Secure Tropos methodology and the systemic perspective of the STPA. This combination enables the representation and analysis of complex and interdependent systems such as those of the C-ES.

4. The SafeSecTropos method

Security and Safety *objectives* describe system features which ensure the system's security and safety. An essential step of the proposed method is the identification of these objectives for each system under analysis. *Constraint* is a restriction related to security/safety objectives which can influence the safety and security analysis and design of a CPS. A security/safety *dependency* introduces security/safety constraint(s) that must be realized to satisfy the corresponding dependency. A security/safety *entity* is a security/safety goal, a security/safety task, or a security/safety resource. A *safety goal* describes the requirement to ensure freedom from accidents/losses. *Safety constraints* are the restrictions in achieving the safety goals. Last, *safety tasks and resources* are the actions needed to achieve the safety goals, and information needed to perform the safety task, respectively. The proposed method integrates the elicitation of security and safety requirements by analyzing the system security and safety constraints in four phases, presented below. In doing so, it encompasses procedural elements of both SecureTropos and STPA. Specifically, it integrates Stage 2 of Secure Tropos with Steps 2 and 3 of STPA, and Stages 3 and 4 of Secure Tropos with Step 4 of STPA, as shown in Figure 1. In particular, during Stage 2, the control diagram has already been developed, and by following Steps 2 and 3 of STPA the safety constraints/objectives for each system are identified. Subsequently, the safety objectives are added to the existing Secure Tropos model, together with the security objectives. Finally, the final architecture of the targeted system along with its detailed design can be developed (Stages 2 and 3 of Secure Tropos), taking into account the causal factors that could lead to an unsafe action (Step 4 of STPA). The integration of the two approaches is depicted in Figure 2.

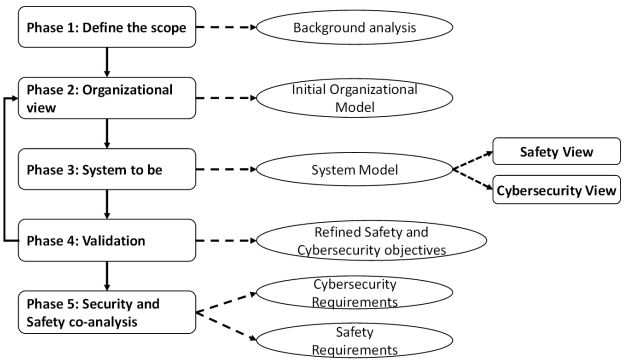


Figure 1: The SafeSecTropos Method

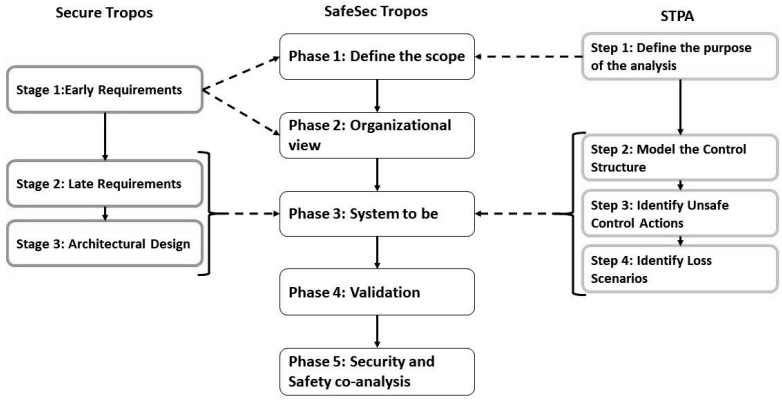


Figure 2: The SafeSecTropos Method

Phase 1 - Define the scope: The scope of the analysis is defined considering both system and environment characteristics. The involved stakeholders are identified, along with the pertinent legislation and standards. Further, their functions and operations are clarified as a step towards the development of the organizational model. The outcome of this phase is the analysis of the targeted ecosystem's background.

Phase 2 - Organizational View: The organizational model of the ecosystem is developed by considering the outcome of Phase 1. In this phase the stakeholders are modelled as actors, and their entities are identified. Such entities can be stakeholder's goals, plans, and resources, along with connections/ interconnections, and dependencies/ interdependencies.

Phase 3 - System-to-be: Modeling and description of the system-to-be. The system under analysis is modelled as an actor and its security and safety objectives are identified. In particular, the system's goals, entities, and processes are identified. Two distinct views are modelled, the safety view and the security view. The former represents system-level hazards and accidents, controller responsibilities, unsafe control actions, causal factors, and safety constraints. The latter depicts system-level vulnerabilities and threats, security entities, goals, and security constraints.

Phase 4 - Validation: The developed models are validated by considering the pertinent legislation, standards, and stakeholders. The outcome is a refined set of safety and security objectives.

Phase 5 - Security and Safety co-analysis: The security analysis for the target system is performed following the SecureTropos process, and the safety analysis following the STPA approach. The outcome of this phase is the set of the security requirements and the set of safety requirements. The prioritization of the security and safety require-

ments should be performed based on specific criteria. These criteria depend on the operational requirements of the system under study, the system architecture, and the validation that relevant stakeholders will perform. According to IEC 63069 [20] the resolution of possible conflicts between safety and security requirements should be performed by relevant stakeholders from both domains. This process is greatly facilitated by the fact that each requirement can be traced back to the objectve(s) that generated it.

5. The case of the C-ES

The CPSs of the C-ES have been selected as the use case of the proposed method because autonomous and remotely controlled vessels are already being extensively developed, and their safety and security requirements have not yet been well studied, in contrast to other domains such as e.g. autonomous vehicles. Furthermore, identifying such requirements is a stepping stone towards designing a secure architecture for such vessels, which will tackle both safety and security issues.

5.1. The C-ES ecosystem

The ICT components and the cyber-physical systems of the C-ES have been identified and analyzed in [26], where also a threat and risk analysis of such systems was carried out, and the most vulnerable onboard CPSs were identified. The most vulnerable onboard systems are the Automatic Identification System (AIS), the Electronic Chart Display System (ECDIS), and the Global Maritime Distress and Safety System (GMDSS) [26]. These are systems responsible for the safe and secure vessel operations.

- The *AIS* is an automated tracking system which facilitates vessel identification, monitoring, and locating. In addition, enhance the collision avoidance capabilities of the vessel. AIS transmits dynamic, voyage, static, and safety data to other vessel systems and to maritime authorities; such data are used to ensure the vessel's safety.
- The *ECDIS* is an information system that supports navigation by providing digital nautical charts, continuously determining the ship's position and unseen hazards. It transmits voyage, dynamic, and static data to facilitate the vessel's voyage and operations.
- The *GMDSS* aims to ensure the availability of the safety-related communication. By leveraging GMDSS the vessel communicates with the shore based station continuously, from any location. GMDSS consists of a set of systems and processes to handle emergencies.

These systems' interconnections, dependencies, and interdependencies were identified in [27] as a step towards eliciting the accordant security requirements in [3].

The stakeholders are the C-ES, the Shore Control Center (SCC), and Other Ships in the vicinity. The identified stakeholders are modelled as actors by leveraging the SecTro Tool. Further, the identification of their goals along with the interconnections, dependencies and interdependencies is performed. As legislation and standards for the autonomous/remotely controlled ships are still under development [28], in our analysis we considered the corresponding ones for conventional ships and systems. The analysis results in the initial organizational model of the targeted ecosystem as depicted in Figure 3. The C-ES includes the Bridge Automation system – BAS, and the Engine Automation System - EAS along with their subsystems.

- An *accident* is the undesired and unplanned event that results in a loss. For the C-ES ecosystem, the accidents [62, 63, 11, 7] are depicted in Table 1.
- A *hazard* is a system state or set of conditions that could lead to an accident (loss). By considering the accident list, the hazards that could lead to these accidents are listed in Table 1.

5.2. Safety and security objectives

The identification of the security and safety objectives facilitates the identification of the security and safety constraints. These objectives are leveraged in modeling the system-to-be towards the elicitation of the system's requirements. The security objectives are based on the *Parkerian Hexad* with the addition of *Non-Repudiation*, which reflects the system nature of the analysis. Although the Parkerian Hexad is based on the CIA model, the added objectives

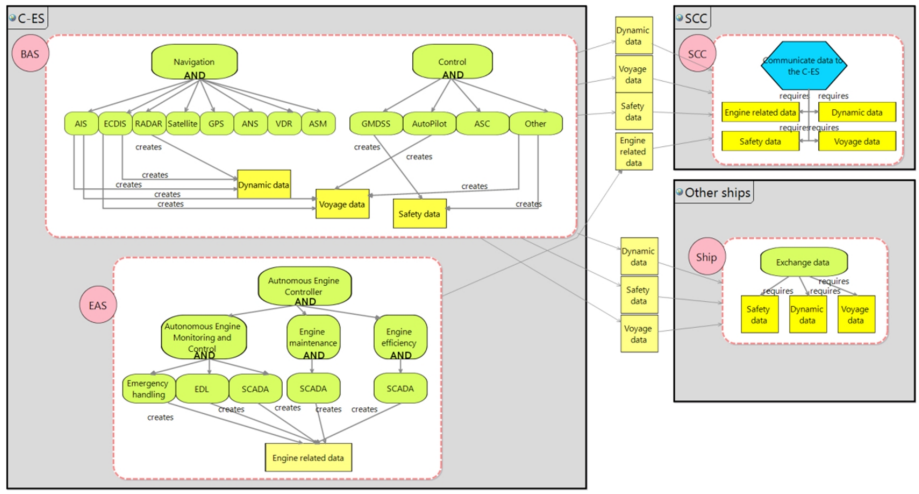


Figure 3: C-ES ecosystem: Organizational view

C-ES Accidents

A	Description
A1	Loss of human life or injury.
A2	Damage in the ship's infrastructure.
A3	Wide energy loss.
A4	Loss of ship's position.
A5	Loss of ship's control.
A6	Loss of the communication links.
A7	Loss of cargo.
A8	Loss of the engine control.
A9	Loss of the control/monitoring of the propulsion/steering system.
A10	Loss of Navigational capabilities.
A11	Loss of ship's stability.
A12	Collision of the vessel with other human made or natural objects.
A13	Fire on board.
A14	Flooding/sinking
A15	Grounding
A16	Environment contaminated.

C-ES Hazards

H	Description	Accidents
H1	Object detection sensor error.	A(1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16)
H2	Software failure.	All
H3	Technical fault (e.g. mechanical fault).	All
H4	Inability to handle harsh weather/sea conditions.	A(1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 16)
H5	Position reference equipment failure.	A(2, 4, 5, 6, 9, 12, 15)
H6	Overloading of the vessel.	A(1, 2, 5, 7, 10, 11, 14, 15, 16)
H7	Shifting of weights.	A(1, 2, 5, 10, 11, 13, 14, 15, 16)
H8	Ignition of electrical equipment or wiring.	All
H9	Passenger starting a fire.	All
H10	Unintended falling overboard.	A(1)
H11	Intended jumping overboard.	A(1)
H12	People getting injured/medical condition.	A(1)

Table 1
C-ES Accidents and Hazards

provide a more comprehensive way to study cybersecurity and data security [49]. Security is better ensured and more efficient countermeasures are designed when all of these six objectives are considered rather than just the CIA triad [54], [46]. These are adapted to fit the maritime domain. All objectives are determined taking into account the operational environment -hence the operational requirements- of each system under study.

- **Confidentiality:** Information exchanged, and communication links between CPSs and services offered by CPSs should be protected against unauthorized access.
- **Integrity:** Information exchanged, services, CPSs, and communication links should be protected against unau-

thorized modifications or manipulations.
- **Availability:** Information exchanged, services, CPSs, and communication links should be available to authorized entities when requested by such entities.
- **Authenticity:** The management, the configuration, and operation of the onboard CPSs and services offered by CPSs should be performed by authorized entities.
- **Possession and Control:** Information exchanged and communication links between CPSs and services offered by CPSs should be protected against the possibility that confidential data be possessed or controlled by unauthorized entities.
- **Utility:** Information exchanged and communication links between CPSs and services offered by CPSs should be useful.
- **Non-Repudiation:** CPSs should not refute responsibility.

Maritime safety is analyzed in [13] as part of transport and safety at sea. Maritime safety aims to protect life, health and property against environmental and operational risks. The safety objectives should ensure that hazards associated with each CPS are identified, tracked, evaluated, and eliminated through the entire system life cycle [24]. Industrial Control Systems attributes described in [32] and performance and security requirements described in [1] are also considered in identifying the maritime safety objectives. The following safety objectives apply.
- **Controllability:** The ability to bring a CPS's/vessel's process into a desired state and handle hazardous events during vessel's operations.
- **Observability:** CPSs should be able to determine their state to enhance the situational awareness of the SCC.
- **Operability:** The CPSs should be able to operate within the constraints imposed by the vessel's state.
- **Resilience:** The CPS's ability to absorb any disturbance caused by faults.
- **Survivability:** The CPS's ability to maintain the vessel's operations at some pre-defined acceptable level.
- **Graceful Degradation:** The CPSs should be able to maintain possibly limited but still safe functionality.
- **Quality of Service:** CPS's data should arrive in time and serve their purpose to perform the necessary safety functions and produce the safety messages that are needed.
- **Availability:** Capability of the CPS to provide a stated function if demanded under given conditions over its defined lifetime.
- **Redundancy:** The systems architecture of the C-ES should be redundant (CPSs, equipment, part, and data redundancy).
- **Fault tolerance:** The CPS of the C-ES should be operational without any interruption from system or software failure.
- **Integrity**: The vessel's CPSs and functions should be durable/stable.

5.3. Applying the SafeSecTropos method to onboard systems

The analysis of the ecosystem in the previous section facilitates the understanding of the systems under study, by providing potential accidents, hazards and the causes of these hazards. The vessel's ecosystem as depicted in Fig. 3 defines the stakeholders along with their interconnections and dependencies with the onboard systems. These remain the same for each individual onboard system under study. The analysis proceeds with the mapping of the identified accidents and hazards to the systems under study; these are used as an input to the STPA analysis. In order to illustrate the workings of the proposed method, we applied it to the AIS, the ECDIS and the GMDSS, these being the most vulnerable onboard systems. The resulting safety and security requirements are presented in Section 5.4. In the interest of saving space, the details of the intermediate steps of the method are presented only for the AIS.

Phase 1: The AIS provides static, dynamic, voyage and safety data, and helps authorities and other ships to monitor sea traffic. The involved stakeholders and their functions and operations are depicted in Fig. 3. Its interactions with other onboard systems and the environment are depicted in Fig. 4.

Phase 2: The organizational view of the AIS is depicted in Fig. 4. The SCC and Other Ships in the vicinity are modelled as system's stakeholders, as in Fig. 3. The onboard subsystems are depicted as separate actors which influence the operation of the AIS, according to their entities (operational and functional requirements). Control flows are exchanged between the AIS, the ECDIS, the Autonomous Navigation System (ANS), and the Collision avoidance system [27]. Therefore, three different control structures result, namely AIS-ANS; AIS-Collision Avoidance; and AIS-ECDIS. Again, in the interest of saving space, only the AIS-ANS control structure is discussed here, shown in Figure 5(a).

Phase 3: In this phase, the AIS goals (green boxes), entities, and resources (yellow boxes) are identified, along with

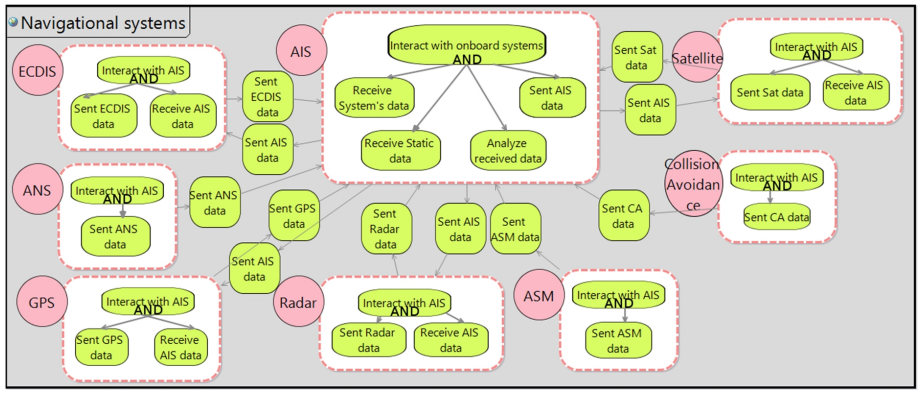
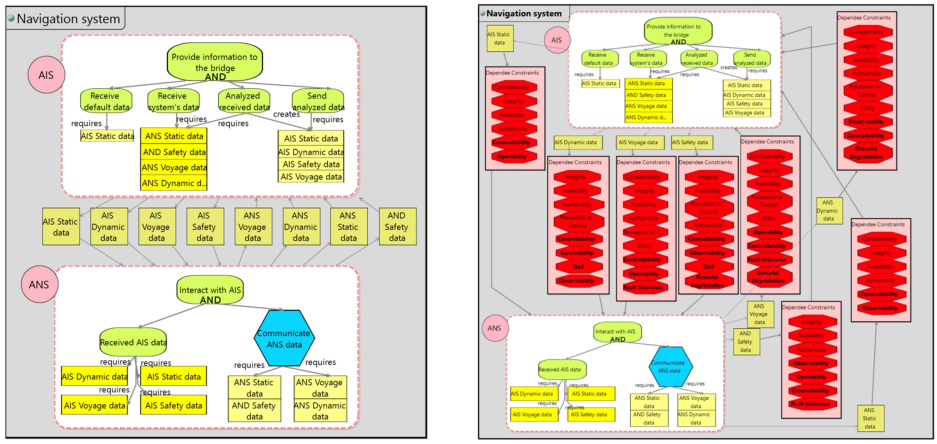


Figure 4: AIS organizational view



(a) Control structure/Safety View (b) security and Safety objectives

Figure 5: AIS and ANS control structures

the security and safety entities (red boxes), by considering Tables 1; these are depicted in Figure 5(a). The security and safety objectives described in Section 5 have been included in the organizational model derived in Phase 2, as shown in Figure 5(b).

Phase 4: The validation of the results derived from Phase 3 (safety and security views) is performed by domain (in our case maritime) experts and relevant stakeholders, who are expected to consider the operational characteristics of autonomous systems, of which the analyst may have limited knowledge. These characteristics are the operational complexity, the environmental complexity, and the system complexity [61]. The former is related to system's deployment and how the system interacts with the surroundings. Environmental complexity captures the complexity of the mission and of the processes of the systems. The latter refers to the functional and operational complexity of the system itself. Both safety and security views should be equally analyzed to extract valid results. Particularly, system level hazards and threats should be identified at the same level of abstraction.

Phase 5: The security analysis of the AIS consists of the identification of the system's vulnerabilities, threats, security objectives, that lead to the identification of the security requirements. The security vulnerabilities for the AIS have been well examined in the literature [18, 6, 10]. Further, G. Kavallieratos et al. in [26] analyzed the security of

AIS Accidents	
A1	Collision of the vessel.
A2	Unable to control the vessel.
A3	Unable to verify the ship's position.
A4	Unable to verify the ship's identity.

AIS Hazards	
H1	Software malfunction/error.
H2	Sensor malfunction/error.
H3	High latency of the transmitting data.
H4	Failure of AIS unit.

Table 2
AIS Accidents and Hazards

C1H1	No patched system.	C1H2	Wrong system's installation.
C2H1	Wrong system configuration.	C2H2	Sensors wrong readings.
C3H1	Lack of maintenance.	C3H2	Lack of sensors redundancy.
C1H3	Improper system configuration.	C1H4	Improper system configuration.
C2H3	Lack of sensors redundancy.	C2H4	Loss of power.
		C3H4	Software error.

Table 3
AIS Hazards Causes

the onboard ICT systems for the C-ES where potential attack scenarios have been developed by using the STRIDE method. Finally, the security objectives, together with the corresponding security requirements for the AIS, have been identified and analyzed in [3]. The STPA principles described in Section 3 are followed towards the safety analysis of the AIS. The identification of the safety goals and constraints requires the identification of AIS accidents and hazards; these are depicted in Table 2. Furthermore, the potential causes of each hazard have been identified in Table 3 to analyze the safety environment of the targeted system and define the corresponding safety objectives.

AIS and ANS control structure: The Control Actions (CA) between the AIS and the ANS are depicted in Fig. 4. These are: CA1: Send AIS dynamic data to ANS; CA2: Send AIS static data to ANS; CA3: Send AIS voyage data to ANS; CA4: Send AIS safety data to ANS. Table 4 depicts the unsafe control actions between the AIS and the ANS, their consequences and the resulting system safety constraints. These lead to the following safety requirements for this control structure. The safety objectives that lead to each requirement follow the requirement in parentheses.

Safety requirements of the AIS-ANS control structure

- SafR1: AIS Dynamic data should be available to the ANS. (Availability, Controllability)
- SafR2: AIS Static data should be available to the ANS. (Availability, Controllability)
- SafR3: Voyage information such as destination port and ETA should be transmitted to the ANS. (Availability, Observability, Controllability, Operability, QoS)
- SafR4: Safety data should be sent to the ANS when needed. (Controllability, Observability, QoS)
- SafR5: The integrity of the transmitted data from the AIS to the ANS and vice versa should be ensured. (Integrity, Controllability, QoS)
- SafR6: Fire alerts should be sent within predefined time limits. (Availability, Survivability, Controllability)
- SafR7: Collision alerts should be transmitted to the ANS within specific time limits. (Availability, Survivability, Controllability)
- Safety alerts should follow a specific structure (eg. Fire/place/time/measures). (Integrity, Controllability, QoS)

5.4. Safety and security requirements of onboard systems

The safety requirements of each onboard system under study are identified by taking into account the safety constraints described in the STPA Tables (such as Table 4) for all relevant control structures. The requirements elicitation is performed by translating the identified safety constraints (e.g. Table 4) into requirements. For example *SafR9* aims to ensure the controllability, redundancy and observability of the AIS services and applications by providing redundancy of the installed sensors. This requirement aims to protect the system against loss of the vessel's control and loss of communication between the AIS and the systems that it interacts with, as a consequence of the Unsafe Control Action 1 - UCA1 depicted in Table 4. These are as follows:

	Control function is not provided	Unsafe control function is provided	Control function is provided in wrong time	Control function is provided for too short or too long
UCA1	AIS dynamic data are not provided to the ANS.	Wrong dynamic data are provided to the ANS.	The AIS dynamic data are provided too soon or too late to the ANS.	Not all AIS dynamic data are provided to the ANS.
UCA2	AIS static data are not provided to the ANS.	Wrong IMO number is provided to the ANS.	AIS static data are provided to the ANS after the entrance to a port.	
UCA3	1) The destination and ETA of the vessel are not provided to the ANS. 2) The ship's draught is not provided to the ANS. 3) The type of the cargo is not provided.	Wrong voyage related data are fed to the ANS.	The AIS voyage data are provided too late to the ANS.	
UCA4	Safety related messages are not sent to the ANS.	1) False fire alert is sent to the ANS. 2) False flooding alert is sent to the ANS. 3) False collision alert is sent to the ANS.	1) Fire alert is sent out of the predetermined time limits. 2) Collision alert is provided to the ANS after the collision.	Fire alert is provided without some details (e.g. missing location).
Conseq.				
CUCA1	1) The ANS is not able to control the vessel. 2) Loss of the communication between the AIS and the ANS. 3) The ANS cannot control the vessel's speed.	1) The ANS suggests the increase/decrease of the vessel's speed. 2) The ANS changes the vessel's heading (misdirection).	1) Loss of ship's position. 2) The ANS is not able to get the navigational status of the vessel.	The ANS cannot continuously communicate with the AIS.
CUCA2	The AIS cannot be authenticated to the ANS.	Insufficient vessel authentication to the ANS.	The ANS cannot control the navigation commands properly.	
CUCA3	The ANS is not able to provide navigational control commands to its sub systems.	Misdirection of the vessel.	1) The ship may enter to no go area. 2) Disruption of vessel's procedures. 3) Vessel's inability to reach port of destination in expected time	
CUCA4	The ANS is not able to provide the necessary functions to address emergencies.	Disruption of vessel's operations.	1) Damage to the ship's infrastructure. 2) Loss of life.	The ANS is not able to send the necessary commands to address the emergencies.

Table 4
AIS to ANS safety constraints

5.4.1. AIS safety requirements

The list below describes the safety requirements of the AIS after the utilization of the SafeSec Tropos. The safety objectives that each requirement fulfills are shown in parenthesis. It can be seen that controllability and obervability are the most prominent objectives for the AIS.

- SafR1: The AIS should be able to send ship's positioning and speed data to the ECDIS. (Controllability, Observability, QoS)
- SafR2: The AIS should be able to send vessel's identification data to the ECDIS. (Controllability, Accessibility, Observability)
- SafR3: The integrity of the charts should be ensured. (Controllability, QoS)
- SafR4: The data sent to ECDIS should be regularly updated. (Operability, Observability, Controllability)
- SafR5: Route and the destination port data of the vessels should be transmitted to the ECDIS. (Observability, Operability, Availability, Controllability)
- SafR6: The necessary AIS data should be provided to the ECDIS to avoid confusion of the system functions. (Observability, Controllability)
- SafR7: The AIS must be patched in case of system vulnerabilities or errors. (Observability, QoS, Controllability, Operability)
- SafR8: The installation and configuration of the AIS must be performed via well trained personnel. (Observability, QoS, Controllability)
- SafR9: The redundancy of the installed sensors should be ensured. (Controllability, Redundancy, Observability)
- SafR10: The power supply of the AIS should be continuous. (Availability, Operability, Controllability)

5.4.2. ECDIS safety requirements

By applying the proposed methodology for the ECDIS, the following safety requirements have been identified by considering the interaction of the targeted system with other onboard CPSs. We notice that availability and QoS are the most prominent safety objectives for the ECDIS.

- SafR1: The redundancy of the installed ECDIS sensors should be ensured. (Redundancy, Controllability, QoS)
- SafR2: Dynamic, Safety, and Voyage data transmitted to the Autonomous Ship Controller (ASC) should be available. (Availability, Observability)
- SafR3: The integrity of the Dynamic, Voyage and Safety data should be ensured. (Integrity, QoS)
- SafR4: Emergency procedures should be initiated when are needed. (Operability, Graceful Degradation)
- SafR5: ECDIS data should be transmitted to the ASC and the ANS in time. (Availability, QoS)
- SafR6: Static data should be provided to the ANS. (Controllability, Operability, Availability)
- SafR7: The authentication of the vessel to the ANS should be ensured. (QoS, Observability)
- SafR8: The integrity of the static data transmitted to ANS should be maintained. (Integrity, QoS)

5.4.3. GMDSS safety requirements

SafeSec Tropos identified the six safety requirements for the GMDSS listed below. Various safety objectives are fulfilled by the identified requirements, since GMDSS is an onboard system that aims to ensure safety during the voyage. As such, different objectives are met by fulfilling the safety requirements.

- SafR1: The availability of the distress signals should be ensured. (Availability, Operability)
- SafR2: The integrity of the transmitted distress signals should be ensured. (Integrity, Observability, Operability)
- SafR3: The Authenticity of the transmitted safety data should be ensured. (Controllability, Operability, Observability)
- SafR4: The redundancy of the communication links between ship to ship and ship to shore should be ensured. (Redundancy, Operability, Availability)
- SafR5: The controllability of the transmitted data and signals should be ensured. (Controllability, Operability, Observability)
- SafR6: The survivability and timeliness of the transmitted safety data should be ensured. (Survivability, Controllability, QoS)

The security requirements for the onboard systems of the C-ES have been identified in [3] by leveraging the Secure Tropos methodology. These are listed below.

5.4.4. AIS security requirements

- SecR1: The AIS should implement the security services in order to protect the system from loss of control or possession of information. (Confidentiality, Authenticity, Possession and Control)
- SecR2: Voyage data such as destination port or cargo related information should be confidential to prevent potential leakage to adversaries. (Confidentiality, Integrity)
- SecR3: The communication channel with the radar system should be redundant. (Availability, Utility)
- SecR4: Voyage related data transmitted to the SCC must be protected against tampering or damage. (Integrity, Availability)
- SecR5: Reliable authentication mechanisms must be in place in order to uniquely identify the actors reading, modifying, and transmitting AIS data, as well as to authenticate the system itself and its services. (Authenticity, Utility, Non-Repudiation)
- SecR6: The AIS must be able to implement lock mechanisms (e.g., lock HMI screen) upon request by the administrator or after a configurable time of idleness. (Confidentiality, Authenticity, Possession and Control, Utility)
- SecR7: The freshness of the dynamic, voyage and safety data should be established. (Availability, Utility)
- SecR8: The configuration and installation of the AIS must be performed by authorized personnel. (Confidentiality, Integrity, Possession and Control, Authenticity)
- SecR9: A suitable amount of AIS sensors should be installed considering the operational mission of the vessel to ensure the redundancy of the AIS. (Availability, Utility)

5.4.5. ECDIS security requirements

- SecR1: The ECDIS administrator must be trained and able to distinguish rogue data packets.(Integrity, Authenticity, Utility)
- SecR2: The use of ECDIS must be restricted only to authorized and well trained personnel. (Confidentiality, Integrity, Authenticity, Possession and Control)
- SecR3: The ECDIS must be able to control the flows of voyage-related data sent to other ships and to the SCC.

(Integrity, Possession and Control, Utility)
- SecR4: The ECDIS should be able to audit sent and received data to external actors. (Integrity, Authenticity, Possession and Control, Utility)
- SecR5: Safety-related information transmitted by the ECDIS must be authenticated. (Integrity, Authenticity, Non-repudiation)
- SecR6: The communication between the ECDIS and the satellite system should be continuously available. (Availability, Utility)

5.4.6. GMDSS security requirements
- SecR1: The authenticity of the transmitted GMDSS signals and data in transit to the ASC, to other subsystems, and to the SCC must be ensured. (Integrity, Authenticity, Non-repudiation)
- SecR2: Distress signals transmitted through the GMDSS must be verified by external actors such as the SCC and other ship's subsystems such as the Autonomous Engine Monitoring and Control (AEMC) and Navigation systems. (Authenticity, Integrity, Utility)
- SecR3: The ASC must be able to provide security, safety, and dynamic data to the GMDSS, when needed. (Availability, Utility)
- SecR4: Safety signals transmitted through the GMDSS to other on board systems and external actors must be continuously available. (Availability, Integrity, Utility)
- SecR5: The GMDSS must be able to detect whether the signal/data comes from a legitimate user/system or from a malicious user. (Confidentiality, Integrity, Authenticity, Non-repudiation)
- SecR6: The signals transmitted to external actors or subsystems must be appropriately encrypted. (Confidentiality, Integrity)
- SecR7: GMDSS antennas must be appropriately installed. (Availability, Utility)

6. Discussion - Challenges, Issues and Observations

Regarding the safety and security requirements derived through the application of SafeSec Tropos to the C-ES case, we note that:

- **Overlapping requirements:** Some safety requirements overlap with security requirements. For example, in the case of the AIS these are: (i) **Saf**R4 with **Sec**R7, (ii) **Saf**R8 with **Sec**R8, and (iii) **Saf**R9 with **Sec**R9. It is noteworthy that similarities can be found in requirements that derive from the availability, integrity, redundancy, and quality of service objectives. Further, the overlapping requirements share the same safety and security goals. The prioritization of the overlapping requirements should be done considering specific criteria described in Section 4.

- **Grouping requirements:** Several requirements, applicable to all systems, can be grouped together to form *Generic requirements*. The remaining requirements will form sets of *System-specific requirements*. This grouping facilitates the communication and the prioritization of the requirements during the architecture definition and implementation phase. Further, the safety and security measures to implement such requirements can follow the same classification; thus their management and implementation process is facilitated.

- **Conflicting requirements:** No obvious requirements conflicts have been identified. This may be attributed to the specificities of the system under study, but also to the careful, non-conflicting, selection of the safety and security objectives and goals that SafeCec Tropos allows.

- **Applicability to C-ES variants:** SafeSec Tropos can capture the differences between different autonomy levels, by modeling the interactions, dependencies and interdependencies of the model components of different vessel types (conventional, remote controlled, autonomous).

- **Validation:** The lack of standards and legislation for autonomous ships and for the joint analysis of safety and security, the validation of safety and security views in Phase 3 of SafeSec Tropos was performed by considering the relevant literature, and security [22] and safety standards [23]. The models derived from phase 3, were validated by considering the operational, environmental, and system complexity of the CPSs under study.

7. Conclusions

In this paper, an integrated approach for safety and security requirements engineering has been proposed. The proposed SafeSec Tropos method facilitates the joint analysis of safety and security by modeling the system for both purposes under the same model and providing documentation regarding the potential conflicts of the identified requirements. These conflicts can be resolved by tracing them back to the corresponding safety and security objectives. Further, complex systems can be analyzed by leveraging the modeling language of the Secure Tropos and the system perspective of the STPA. The safety and security objectives in the maritime domain have been identified towards the identification of cyber physical systems safety and security requirements. Due to the complex nature of such systems, a graphical-based model is proposed by leveraging the existing SecTro tool. The ecosystem of the Cyber-Enabled Ship is used to demonstrate the applicability of the proposed method. The three most vulnerable onboard systems, namely the AIS, the ECDIS, and the GMDSS have been analyzed and their safety and security requirements have been identified.

As future work we plan to define a safety and security architecture compliant to the identified requirements. Such an architecture would possibly include automated incident response mechanisms [36] to handle the critical incidents that may occur in the vessel's infrastructure. The identification and modelling of the appropriate security and safety measures will facilitate the design and the installation of safe and secure by design systems in critical infrastructures such as the C-ES. Furthermore, it is important to study and explore how potential changes in one set of requirements affects both sets and how this can be done in an iterative fashion. Additionally, in our future work we plan to verify the applicability and usefulness of the SafeSec Tropos along with its claimed advantages in other reference architectures and domains where the emerging technology of the cyber-physical systems also exists. Such domains could be autonomous vehicles, smart homes [15], and various fields within Industry 4.0 [41]. Finally, due to the increasing development of CPSs for the maritime domain to facilitate port, vessel, and many logistics operations, it is important to make concrete proposals for creating a standard for safety and security in the maritime domain, similar to e.g. IEC TR63069:2019 [20].

References

[1] C. Alcaraz and J. Lopez. Analysis of requirements for critical control systems. *International journal of critical infrastructure protection*, 5(3-4):137–145, 2012.

[2] S. Ali, T. Al Balushi, Z. Nadir, and O. K. Hussain. Standards for CPS. In *Cyber Security for Cyber Physical Systems*, pages 161–174. Springer, 2018.

[3] anonymized. under revision. 2019.

[4] P. Asare, J. Lach, and J. A. Stankovic. FSTPA-I: A formal approach to hazard identification via system theoretic process analysis. In *Proceedings of the ACM/IEEE 4th International Conference on Cyber-Physical Systems*, pages 150–159. ACM, 2013.

[5] F. Asplund, J. McDermid, R. Oates, and J. Roberts. Rapid integration of CPS security and safety. *IEEE Embedded Systems Letters*, 11(4):111–114, 2019.

[6] M. Balduzzi. AIS Exposed Understanding Vulnerabilities Attacks 2.0. In *Blackhat, Asia, 2014*, page 44, 2014.

[7] O. A. V. Banda and S. Kannos. Hazard analysis process for autonomous vessels. Technical report, 2017.

[8] V. Bolbot, G. Theotokatos, E. Boulougouris, and D. Vassalos. Safety related cyber-attacks identification and assessment for autonomous inland ships. In *International Seminar on Safety and Security of Autonomous Vessels (ISSAV)*, September 2019.

[9] V. Bolbot, G. Theotokatos, L. M. Bujorianu, E. Boulougouris, and D. Vassalos. Vulnerabilities and safety assurance methods in Cyber-Physical Systems: A comprehensive review. *Reliability Engineering System Safety*, 182:179 – 193, 2019.

[10] D. Bothur, G. Zheng, and C. Valli. A critical analysis of security vulnerabilities and countermeasures in a smart ship system. *In Proceedings of 15th Australian Information Security Management Conference, Edith Cowan University, Perth, Western Australia. (pp.81-87)*, 2017.

[11] DNV-GL. Autonomous and remotely operated ships, class guideline. Technical report, 2018.

[12] D. P. Eames and J. Moffett. The integration of safety and security requirements. In *International Conference on Computer Safety, Reliability, and Security*, pages 468–480. Springer, 1999.

[13] K. Formela, A. Weintrit, and T. Neumann. Overview of definitions of maritime safety, safety at sea, navigational safety and safety in general. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 13, 2019.

[14] I. Friedberg, K. McLaughlin, P. Smith, D. Laverty, and S. Sezer. STPA-SafeSec: Safety and security analysis for cyber-physical systems. *Journal of Information Security and Applications*, 34:183 – 196, 2017.

[15] K. Ghirardello, C. Maple, D. Ng, and P. Kearney. Cyber security of smart homes: Development of a reference architecture for attack surface analysis. In *Living in the Internet of Things: Cybersecurity of the IoT - 2018*, pages 1–10, March 2018.

[16] J. Glomsrud and J. Xie. A Structured STPA Safety and Security Co-analysis Framework for Autonomous Ships. *Safety and Reliability–Safe Societies in a Changing World. Proceedings of ESREL 2018, June 17-21, 2018, Trondheim, Norway*, 2019.

[17] N. H. C. Guzman, D. K. M. Kufoalor, I. Kozin, and M. A. Lundteigen. Combined safety and security risk analysis using the UFoI-E method: A case study of an autonomous surface vessel. In *29th European Safety and Reliability Conference*, pages 4099–4106, 2019.

[18] J. Hall, J. Lee, J. Benin, C. Armstrong, and H. Owen. IEEE 1609 Influenced Automatic Identification System (AIS). In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5, 2015.
[19] International Electrotechnical Commission. Functional safety - Safety instrumented systems for the process industry sector, IEC 61511 . Technical report, 2016.
[20] International Electrotechnical Commission. Industrial-process measurement control and automation, Framework for functional safety and security IEC 63069:2019. Technical report, 2019.
[21] International Maritime Organization . IMO takes first steps to address autonomous ships. `http://www.imo.org/en/mediacentre/pressbriefings/pages/08-msc-99-mass-scoping.aspx`, 2018.
[22] International Organization for Standardization (ISO). Information technology — Security techniques — Information security management systems, ISO 27000 series. Technical report, 2016.
[23] International Organization for Standardization (ISO). Road vehicles — Functional safety, ISO 26262-1:2018. Technical report, 2018.
[24] Joint Software System Safety Committee. Software system safety handbook, a technical managerial team approach. Technical report, 1999.
[25] N. Karanikas. Revisiting the relationship between safety and security. *International journal of safety and security engineering*, 8(4):547–551, 2018.
[26] G. Kavallieratos, S. Katsikas, and V. Gkioulos. Cyber-attacks against the autonomous ship. In *Computer Security*, pages 20–36, Cham, 2019. Springer International Publishing.
[27] G. Kavallieratos, S. Katsikas, and V. Gkioulos. Modelling shipping 4.0: A reference architecture for the cyber-enabled ship. In *Proceedings of the 12th Asian Conference on Intelligent Information and Database Systems*, 2020.
[28] A. Komianos. The autonomous shipping era. operational, regulatory, and quality challenges. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 12, 2018.
[29] L. Kretschmann, Ø. J. Rødseth, B. S. Fuller, H. Noble, J. Horahan, and H. McDowell. MUNIN D9.3: Quantitative assessment. Technical report, 2015.
[30] L. Kretschmann, Ø. J. Rødseth, A. Tjora, B. S. Fuller, H. Noble, and J. Horahan. MUNIN D9.2: Qualitative assessment. Technical report, 2015.
[31] S. Kriaa, L. Pietre-Cambacedes, M. Bouissou, and Y. Halgand. A survey of approaches combining safety and security for industrial control systems. *Reliability engineering & system safety*, 139:156–178, 2015.
[32] M. Krotofil, K. Kursawe, and D. Gollmann. Securing industrial control systems. In *Security and Privacy Trends in the Industrial Internet of Things*, pages 3–27. 2019.
[33] N. Leveson. *Engineering a safer world: Systems thinking applied to safety*. MIT press, 2011.
[34] N. Leveson and J. Thomas. STPA handbook. 2018.
[35] E. Lisova, I. Šljivo, and A. Čaušević. Safety and security co-analyses: A systematic literature review. *IEEE Systems Journal*, 13(3):2189–2200, Sep. 2019.
[36] J. Lopez, C. Alcaraz, and R. Roman. Smart control of operational threats in control substations. *Computers & Security*, 38:14–27, 2013.
[37] M. A. Lundteigen and B. A. Gran. The need of improved methods to handle functional safety and cybersecurity in industrial control and safety systems.
[38] X. Lyu, Y. Ding, and S.-H. Yang. Safety and security risk assessment in cyber-physical systems. *IET Cyber-Physical Systems: Theory & Applications*, 2019.
[39] N. R. Mead. How to compare the security quality requirements engineering (square) method with other methods. Technical report, Carnegie-Mellon University, 2007.
[40] D. Mellado, C. Blanco, L. E. Sánchez, and E. Fernández-Medina. A systematic review of security requirements engineering. *Computer Standards & Interfaces*, 32(4):153–165, 2010.
[41] H. Mouratidis and V. Diamantopoulou. A security analysis method for industrial internet of things. *IEEE Transactions on Industrial Informatics*, 14(9):4093–4100, 2018.
[42] H. Mouratidis and P. Giorgini. Secure tropos: a security-oriented extension of the tropos methodology. *International Journal of Software Engineering and Knowledge Engineering*, 17(02):285–309, 2007.
[43] D. Muñante, V. Chiprianov, L. Gallon, and P. Aniorté. A review of security requirements engineering methods with respect to risk analysis and model-driven engineering. In *International Conference on Availability, Reliability, and Security*, pages 79–93. Springer, 2014.
[44] A. Nhlabatsi, B. Nuseibeh, and Y. Yu. Security requirements engineering for evolving software systems: A survey. In *Security-aware systems applications and software development methods*, pages 108–128. IGI Global, 2012.
[45] T. Novak and A. Treytl. Functional safety and system security in automation systems-a life cycle model. In *2008 IEEE International Conference on Emerging Technologies and Factory Automation*, pages 311–318. IEEE, 2008.
[46] D. B. Parker. Toward a new framework for information security? *Computer security handbook*, pages 3.1 – 3.23, 2012.
[47] S. Paul and L. Rioux. Recommendations for security and safety co-engineering (release n 2). *WIT Transactions on The Built Environment*, 151:335 – 349, 2015.
[48] M. Pavlidis, S. Islam, and H. Mouratidis. A case tool to support automated modelling and analysis of security requirements, based on secure tropos. In *International Conference on Advanced Information Systems Engineering*, pages 95–109. Springer, 2011.
[49] G. Pender-Bey. The parkerian hexad, master's thesis. Technical report, 2016.
[50] D. Pereira, C. Hirata, R. Pagliares, and S. Nadjm-Tehrani. Towards combined safety and security constraints analysis. In S. Tonetta, E. Schoitsch, and F. Bitsch, editors, *Computer Safety, Reliability, and Security*, pages 70–80, Cham, 2017. Springer International Publishing.
[51] L. Pièrre-Cambacédès and M. Bouissou. Modeling safety and security interdependencies with bdmp (boolean logic driven markov processes). In *2010 IEEE International Conference on Systems, Man and Cybernetics*, pages 2852–2861. IEEE, 2010.
[52] N. Polatidis, M. Pavlidis, and H. Mouratidis. Cyber-attack path discovery in a dynamic supply chain maritime risk management system. *Computer Standards & Interfaces*, 56:74–82, 2018.

[53] C. Raspotnig and A. Opdahl. Comparing risk identification techniques for safety and security requirements. *Journal of Systems and Software*, 86(4):1124 – 1151, 2013. SI : Software Engineering in Brazil: Retrospective and Prospective Views.

[54] R. C. Reid and A. H. Gilbert. Using the parkerian hexad to introduce security in an information literacy class. In *2010 Information Security Curriculum Development Conference*, pages 45–47, 2010.

[55] S. Sadvandi, N. Chapon, and L. Pietre-Cambacédes. Safety and security interdependencies in complex systems and SoS: Challenges and perspectives. In *Complex Systems Design & Management*, pages 229–241. Springer, 2012.

[56] M. Z. Schmittner, C. and P. Puschner. Limitation and improvement of STPA-Sec for safety and security co-analysis. In *Proceedings of the SAFECOMP 2016 Workshops*, LNCS 9923, pages 195 —- 209. Springer, 2016.

[57] L. Shan, B. Sangchoolie, P. Folkesson, J. Vinter, E. Schoitsch, and C. Loiseaux. A survey on the applicability of safety, security and privacy standards in developing dependable systems. In *International Conference on Computer Safety, Reliability, and Security*, pages 74–86. Springer, 2019.

[58] S. M. Sulaman, A. Beer, M. Felderer, and M. Höst. Comparison of the fmea and stpa safety analysis methods–a case study. *Software Quality Journal*, 27(1):349–387, 2019.

[59] M. Sun, S. Mohan, L. Sha, and C. Gunter. Addressing safety and security contradictions in cyber-physical systems. In *Proceedings of the 1st Workshop on Future Directions in Cyber-Physical Systems Security (CPSSW'09)*. Citeseer, 2009.

[60] E. N. Torkildson, J. Li, S. O. Johnsen, and J. A. Glomsrud. Empirical studies of methods for safety and security co-analysis of autonomous boat. *Safety and Reliability–Safe Societies in a Changing World. Proceedings of ESREL 2018, June 17-21, 2018, Trondheim, Norway*, 2018.

[61] I. B. Utne, A. J. Sørensen, and I. Schjølberg. Risk management of autonomous marine systems and operations. In *ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering*. American Society of Mechanical Engineers Digital Collection, 2017.

[62] K. Wróbel, J. Montewka, and P. Kujala. System-theoretic approach to safety of remotely-controlled merchant vessel. *Ocean Engineering*, 152:334–345, 2018.

[63] K. Wróbel, J. Montewka, and P. Kujala. Towards the development of a system-theoretic model for safety assessment of autonomous merchant vessels. *Reliability Engineering & System Safety*, 178:209–224, 2018.

[64] W. Young and N. Leveson. Systems thinking for safety and security. In *Proceedings of the 29th Annual Computer Security Applications Conference*, ACSAC '13, pages 1–8, New York, NY, USA, 2013. ACM.

10 Article VI: Attack Path Analysis for Cyber Physical Systems [6]

Attack Path Analysis for Cyber Physical Systems

CyberICPS 2020, SECPRE 2020, ADIoT 2020: Computer Security pp 19-33, Springer, (LNCS, volume 12501)

Georgios Kavallieratos[1][0000 0003 1278 1943] and Sokratis Katsikas[1,2][0000 0003 2966 9683]

[1] Norwegian University of Science and Technology, Department of Information Security and Communications Technology, Gjøvik, Norway
(name.surname)@ntnu.no

[2] Open University of Cyprus, School of Pure and Applied Sciences, Latsia, Nicosia, Cyprus
sokratis.katsikas@ouc.ac.cy

Abstract. The identification and analysis of potential paths that an adversary may exploit to attack Cyber Physical Systems comprising subsystems enables the comprehensive understanding of the attacks and the impact that may have to the overall system, thus facilitating the definition of appropriate countermeasures that will satisfy the pertinent security requirements. To this end, several attack modelling techniques can be employed, the attack graph being the most prevalent among them. Unfortunately, the discovery and analysis of all possible attack paths in an attack graph is not possible in systems even of a moderate size. In this work we propose a novel systematic method for discovering and analyzing attack paths in real-world scale interconnected Cyber Physical Systems. The method considers the criticality of each sub-system in discovering paths and the risk to the overall system that each path presents to analyze and prioritize paths. We illustrate the workings of the method by applying to the navigational Cyber Physical Systems of the Cyber-Enabled Ship to identify and analyze highly critical attack paths originating from the Automatic Identification System (AIS) and targeting the Autonomous Navigation System (ANS).

Keywords: Cyber Physical Systems · Attack path analysis · Navigational system · Autonomous ships.

1 Introduction

Various cyberattacks targeting Cyber Physical Systems (CPSs) have been reported and analyzed in the last decade [1]. Such attacks may have severe impact on both the physical and the cyber parts of the CPS. This is particularly so in autonomous systems, as the higher the level of autonomy, the greater the impact

of a cyberattack, due to the extended interconnections and interdependencies among the networked components of such systems [2].

The fourth industrial revolution in shipping is known as cyber-shipping or Shipping 4.0 [3]. This digital transformation increases the cyber risks in the already vulnerable to cyberattacks maritime domain. Various cyberattacks in this domain have occurred, have been studied and analyzed in the literature [4,5,6]; the increasing proliferation of interconnected on-board CPSs increases the attack surface of contemporary vessels. The emerging technology of the remotely controlled and autonomous vessels, both variants of the Cyber-Enabled Ship (C-ES) [7], will increase even further the attack surface. Thus, C-ESs of the future will need to be cyber-secure-by-design. The analysis of potential cyberattacks that target CPSs of the C-ES is an important step in this process, as it provides comprehensive insight into possible attacks and facilitates the identification of the necessary mitigation strategies and measures.

Attack models are an important instrument for improving our perception and understanding of cyberattacks; both are fundamental in evaluating the security of a networked system and in subsequently selecting appropriate countermeasures [8]. Attack models are the result of employing attack modelling techniques, that allow the representation of the sequence of events that lead to a successful cyberattack. Such techniques are grouped in three categories, namely (1) techniques that are based on the use case framework; (2) techniques that present a cyberattack from a temporal perspective; and (3) graph based techniques [9]. Among the latter, attack graphs and attack trees are the most commonly used methods for representing cyberattacks.

Attack graphs are conceptual diagrams used to analyze how a target can be attacked, so as to improve its security posture. This is performed in four stages, namely (1) Acquisition of system information; (2) Attack graph generation; (3) Attack graph analysis; and (4) Use of the results. In the first stage, information about the system (e.g. network topology, sub-systems, vulnerabilities, network configuration, connectivity) is collected. This information is subsequently used to generate the attack graph, which is then used for performing the analysis of attacks. Finally, the results of the analysis are used to inform the risk management process.

In a system of networked assets, whereby an asset may well be a system in its own right, an *attack path* is an ordered sequence of assets that can be used as stepping stones by an attacker seeking to attack one or more assets on the path.

The main advantage of an attack graph over other types of attack models is that it helps to identify all possible attacks on a system [10]. Notwithstanding the advantages of graph-based attack models in describing important elements of a cyberattack, these models suffer from a scalability problem if all possible attack paths are considered [11]. This is why, even though the analysis of all attack paths can lead to the identification of the optimal security solutions, techniques that allow the identification of those attack paths that present the most significant risk to the overall system are sought. Examples of such techniques are [12,13,14].

The analysis of potential attack paths is commonly based only on the *vulnerabilities* of the systems on the attack path. This limits considerably the insight into the possible attack scenarios, and limits the subsequent selection of countermeasures to only those that reduce the vulnerability, excluding countermeasures that reduce the other elements of risk, namely the likelihood of the threats and the extent of the impact, and their combinations.

In this paper we propose a method for cyberattack path discovery and prioritization for CPSs comprising a number of sub-systems. The method is based on the criticality of the sub-systems on each path and on the cyber risk to the overall system that each attack path represents. Thus, we provide a holistic view of the attack, that can be further exploited in designing the necessary and most appropriate mitigation techniques and strategies.

The most vulnerable CPSs on board the C-ES are those comprising the navigational system [7]. We therefore illustrate the workings of our method by applying it to the navigational CPS system of the C-ES.

The contribution of this work is twofold:

- We have developed a novel method for discovering and analyzing attack paths in interconnected CPSs, and
- we have applied it to discover and analyze attack paths for the navigational CPSs in a C-ES.

The remaining of the paper is structured as follows: Section 2 reviews the related work. Section 3 describes the proposed method. In section 4 the method is applied to the navigational system of the C-ES. Finally, section 5 summarizes our findings and indicates possible future work.

2 Related work

Attack graphs find their origins in Dacier's PhD thesis and early papers [15],[16], [17], where the concept of the *privilege graph* was introduced. The concept of the *attack graph* was proposed in [18]. Attack graphs are classified into five categories, namely *generic*; *alert correlation*; *vulnerability*; *miscellaneous*; and *dependency* [9]. Several approaches for attack graph generation and analysis have been proposed in the literature. S. Khaitan et al. in [19] surveyed approaches that generate attack graphs in wired and wireless networks, and focused on the limitation of existing approaches to handle complex and scalable networks. Typically, graph construction attempts to identify all possible attacks paths [20]. The process may also be supported by software tools, such as the early tool presented in [21], MulVal [22]; TVA [23]; NuSMV [24]. A survey of attack graph analysis methods can be found in [25].

According to [26], attack graphs face a combinatorial explosion. Thus they can be applied to small network systems only [23]; for large-scale systems it is necessary to reduce the complexity of the attack graph. Methods for doing so include path pruning, network properties compression, and property matching

time reducing [25]. Examples of such methods are found in [24], where a Breadth-first search method is used to identify the vulnerabilities and build the attack graph; in [27], that introduces the concept of group reachability to reduce graph complexity; in [28], where the authors propose a multi-agent-based distributed approach to generate the attack graph using Depth-first search; in [29], where the use of a dynamic algorithm that generates an attack graph consisting of the K most probable to be exploited attack paths; in [30], where a a Bayesian-based attack graph generation method is proposed; in [31] that is based on a cut and divide method and a series of division rounds and uses Depth-first search to search the smaller graphs; and in [32], where the authors exploit risk flow within an attack graph for performing security risk assessment. J. H. Castellanos et al. in [33] propose a method to identify attack paths that uses data-flow graphs, and N. Polatidis et al. in [34] propose an attack path discovery method that is used as a component of a maritime risk management system. The method uses constraints and Depth-first search to effectively generate attack graphs and has been used for identifying attack paths and security mechanisms in the maritime domain [34], [35].

The main characteristics and goals of attack graph analysis methods for CPSs have been discussed in [36]. Out of the nine methods examined therein, only one considers potential security risks in analyzing and prioritizing potential attack paths, whilst the rest focus on vulnerabilities for performing this analysis; this is also the case with all the methods referenced above.

In the C-ES context, safety-related cyberattacks for autonomous inland ships have been studied in [37]. Cyberattack scenarios against autonomous ships have been analyzed in [7] by leveraging the STRIDE methodology. However, none of these works considered possible paths that an attacker may follow to launch a cyberattack against a C-ES.

3 Discovering and analyzing attack paths

3.1 Problem formulation

We assume a CPS comprising sub-systems that is described by a directed graph *G(V,E)* whose nodes respresent the sub-systems and the edges represent interconnections between nodes. The goal is to discover and analyze attack paths between selected *entry* and *target* sub-systems, based on information regarding the criticality of the sub-systems and the overall cyber risk to the overall system that an attack path represents. The results are to be used to inform the risk management process in selecting appropriate countermeasures to reduce the overall cyber risk.

3.2 Components of the proposed method

The proposed method integrates a number of components, that are briefly described in this section.

Identifying critical components in CPSs: Because of the distributed nature of almost all CPSs, in many cases suffices to destroy or damage only a few influential nodes or links in a system to inflict failure of the entire system. An aggregated index (the Z index) that leverages the characteristics of both nodes and links to rank the components of a CPS according to their criticality, and a method to calculate it by means of a multiple attribute decision making (MADM) method was proposed in [38]. The method involves the use of novel graph metrics, namely the *Tacit Input Centrality (TIC)* and the *Tacit Output Centrality (TOC)* that measure how frequently each link in a system is utilized and reflect the importance of a link in relation to the nodes it connects. It also involves the *Closeness Centrality (CC)* of a node that measures how close the node is to all other nodes, by calculating the shortest path length from the node to every other node in the network.

Estimating the risk of each CPS component: The DREAD method was developed by Microsoft as a complement to STRIDE [39], to provide a quantitative estimate of the risk in a software system [40]. DREAD stands for *Damage*, *Reproducibility*, *Exploitability*, *Affected users*, and *Discoverability*. *Damage* represents the damage that a cyber-attack may inflict to the system; together with *Affected Users/Systems* they reflect the *Impact* of the attack. *Reproducibility* reflects the ability of the attacker to reproduce the attack, and *Exploitability* represents the ability to exploit the system's vulnerabilities and perform the attack. *Discoverability* reflects the capacity of the adversary to identify system vulnerabilities. The sum of *Reproducibility*, *Exploitability*, and *Discoverability* reflects the *Likelihood* of the cyberattack [41].

Each of the DREAD variables accepts an integer value in [0,3], the value being assigned by considering the criteria listed in Table 1 that is adapted from [40] to capture also aspects of CPSs.

	High (3)	**Medium (2)**	**Low (1)**
D	The adversary is able to bypass security mechanisms; get administrator access; upload/modify the CPS content.	Leakage of confidential information of the CPS (functions/source code); inflict partial malfunction/disruption to the system.	Leaking non-sensitive information; the attack is not possible to be extended over other CPSs.
R	The cyberattack can be reproduced anytime to the targeted CPS.	The adversary is able to reproduce the attack but under specific risk conditions.	Although they know CPS's vulnerabilities/faults, the attacker is not able to perform the cyberattack.
E	The cyberattack can be performed by a novice adversary in a short time.	A skilled adversary could launch the attack.	The attack requires an extremely skilled person and in-depth knowledge of the targeted CPS.

A	All CPSs are affected	Partial users/systems, non-default configuration	The attack affects only the targeted CPS.
D	The CPS's vulnerabilities are well known and the attacker is able to get access to the relevant information to exploit the vulnerabilities.	The CPS's vulnerabilities/faults are not well known and the adversary needs to get access to the CPS.	The threat has been identified and the vulnerabilities have been patched.

Table 1: DREAD Criteria

The DREAD score is calculated as follows [41]:

$$\frac{\sum(Damage, Affectedsystems)}{2} = Impact \tag{1}$$

$$\frac{\sum(Reproducibility, Exploitability, Discoverability)}{3} = Likelihood \tag{2}$$

$$DREADscore = \frac{(Impact + Likelihood)}{2} \tag{3}$$

The DREAD risk level is determined as follows:

- **If** *DREAD score* ≤ 1 **then** *DREAD risk level* := Low
- **If** $1 <$ *DREAD score* ≤ 2 **then** *DREAD risk level* := Medium
- **If**: $2 <$ *DREAD score* ≤ 3 **then** *DREAD risk level* := High

Integrating the stakeholders' views: The assessment of the importance of each possible attack path is based upon the combination of two values, namely the risk of each CPS component on the path (as estimated by e.g. the DREAD method); and the effect that a failure of each such component would have to the operation of the overall system, as seen from the perspective of the system stakeholders; this is captured by the *CPSImp* metric. *CPSImp* is assigned to each CPS by the administrator/designer/operator/relevant stakeholder of the system to reflect the importance of each sub-system to the overall system. It can take one of three distinct values as follows:

- 1: Low importance (potential system damage or disruption cannot inflict any significant damage to the overall system);
- 2: Medium importance (if the system is damaged or disrupted, overall system malfunctions may occur, but no crucial deviation from normal operation);
- 3: High importance (if the system is damaged or disrupted, the operation of the overall system will be severely affected).

The importance of the overall attack path taking into account both the risk level and the stakeholders' view is calculated according to the following equation [35]:

$$AttackPathImportance = 0.6 * CPSImp + 0.4 * Risk \quad (4)$$

3.3 Input data

The proposed method operates on the following input data:

1. A directed graph *G(V,E)* representing the CPS under study, as defined in section 3.1. Such a graph can be generated using automated tools such as the CASOS ORA tool from Carnegie Mellon University [42].
2. The entry CPS (e) and the targeted CPS (t) in G.
3. The profile of the assumed adversary. One of the novel features of the proposed approach is that it is both risk driven (as opposed to only vulnerability driven) and is intended to in turn drive the subsequent risk management process. Thus, the adversary model must also be considered when discovering and analyzing attack paths, following the suggestion in NIST SP800-30 [43]. The adversary is profiled by means of the following attributes, adapted from [35]:
   - *Accessibility* is a measure of the adversary's logical and physical accessibility of the adversary to the attack surface of each entry sub-system. It assumes a "yes" or "no" value.
   - *Capability* represents the ability of the adversary to access the necessary resources (technical, physical, and logical) to perform an attack against each entry sub-system. It is measured in a qualitative scale ranging from "Low "to "Medium "to "High ".
   - *Motivation* represents the determination of the adversary to carry out the attack. It is measured in a qualitative scale ranging from "Low "to "Medium "to "High ".

   When the adversary does not have the required levels of accessibility, capability, and motivation, there are no possible attack paths.

3.4 The proposed method

As shown in Figure 1 the proposed method is structured in six steps. These are described below.

1. **Step 1 - Load input data:** All input data as specified in Figure 1 are loaded.
2. **Step 2 - Check adversary profile:** The profile of the adversary is checked against threshold values. If the adversary is deemed incapable of launching an attack against e, no possible attack paths exist and the method terminates.

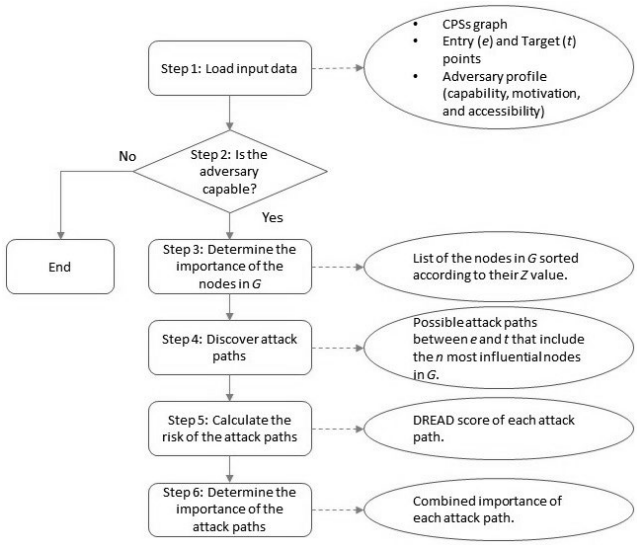


Fig. 1: Process

3. **Step 3 - Determine the criticality of the nodes in G:** The method in [38] is applied to G to determine the criticality of each node. The result of this step is a list of all nodes in G sorted according to their Z value in ascending order[3] (the L list). The reader is referred to [38] for the detailed workings of the method that are hereby omitted in the interest of saving space.

4. **Step 4 - Discover attack paths:** By performing a depth-first search, all non-circular paths starting at e and terminating at t that include at least one of the top n nodes in L are discovered.

5. **Step 5 - Calculate the risk of the attack paths:** The risk to the overall system that each attack path among those discovered in Step 4 represents is calculated, by applying the DREAD method [39] on each of the nodes on each path. The risk of the path equals the maximum risk of its nodes.

6. **Step 6 - Determine the importance of the attack paths:** The importance of each attack path among those discovered in Step 4 is calculated by means of equation (4), and the list of attack paths is prioritized.

[3] The lower the Z value of a sub-system the more critical the sub-system is.

3.5 Characteristics of the method

The proposed method enjoys some desirable characteristics that are not always shared with alternative methods for discovering and analyzing attack paths in CPSs:

- The proposed method allows the analysis of attack paths against *composite CPSs* i.e. cyber-physical systems that comprise subsystems; it thus constitutes a step towards attack path analysis against systems-of-systems.
- The proposed method incorporates a component that allows the *identification of critical subsystems* in a composite CPS. This is particularly useful when designing the set of countermeasures, as the protection of critical subsystems would be prioritized.
- The proposed method analyzes attack paths by considering *all the elements of risk* rather than simply vulnerabilities. This is also particularly useful when designing the set of countermeasures, as it allows the informed selection of controls that may reduce more than one of the elements of risk.
- The proposed method incorporates a component that *involves the stakeholders* to determine the importance of the discovered attack paths, thus enabling the extraction of realistic results, particularly in complex environments where multiple stakeholders exist.
- The proposed method *scales well* with the number of subsystems of the composite CPS.
- The proposed method is *domain-agnostic*; it can be applied in any CPS domain.

4 Attacks against the navigational CPSs of the C-ES

The generic ICT architecture of the Cyber-Enabled Ship in the form of a hierarchical tree structure was proposed in [7]. The detailed interconnections, dependencies and interdependencies among the CPSs of the C-ES, including those in the navigational system were determined in [2]. The latter, along with their interconnections are depicted in Figure 2. According to [7], [44], the three most vulnerable systems on board the C-ES are the Automatic Identification System (AIS), the Electronic Chart Display Information System (ECDIS), and the Global Maritime Distress and Safety System (GMDSS); among these, the Automatic Identification System (AIS) is the most vulnerable. On the other hand, a potential failure of the Autonomous Navigation System (ANS) or the Autonomous Ship Controller (ASC) can result in a cascade failure effect among the CPSs of the C-ES, with significant impact [2]. Accordingly, in order to illustrate the workings of the proposed method, we selected to analyze attack paths for the navigational system of the C-ES that have as entry point the AIS and as target system the ANS.

Assuming that the adversary is deemed capable of launching the attack, Step 3 of the proposed method returns the *Tacit Input Centrality - TIC*, *Tacit Output*

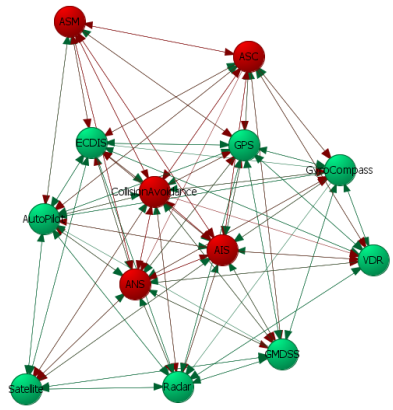


Fig. 2: Navigational CPSs of the C-ES

Table 2: Navigational CPSs metrics

	ANS	**AIS**	ECDIS	RADAR	GPS	**ASC**	**C.A.**	**ASM**	AP	VDR	Gyro	GMDSS	Satellite
TIC	0.772	0.590	0.545	0.409	0.50	1	0.590	0.636	0.545	0.545	0.409	0.590	0.5
TOC	0.727	0.590	0.545	0.409	0.50	1	0.590	0.636	0.545	0.545	0.409	0.181	0.045
CC	0.767	0.697	0.676	0.657	0.657	0.920	0.719	0.719	0.697	0.657	0.622	0.697	0.657
Z	**0.538**	**0.851**	0.940	1.193	1.116	**0**	**0.843**	**0.736**	0.933	0.803	0.984	1.034	1.53

Centrality - TOC, *Closness Centrality - CC*, and *Aggregated index - Z* values for the systems in Figure 2 as depicted in Table 2.

Assuming that we are interested in analyzing attack paths that include the five most critical components, we set n equal to 5 in Step 4 of the proposed method. The five systems with the lowest Z values are shown as red nodes in Figure 2. Step 4 then results in identifying sixteen attack paths having as entry system the AIS and as target system the ANS. These are depicted in Table 3.

Table 3: Attack paths from AIS to ANS

Path ID	Cyber-attack path
1	AIS, ANS
2	AIS, ASC, ANS
3	AIS, ASC, ASM, ANS
4	AIS, ASC, ASM, C.A., ANS
5	AIS, ASC, C.A., ANS
6	AIS, ASC, C.A., ASM, ANS
7	AIS, ASM, ANS
8	AIS, ASM, ASC, ANS
9	AIS, ASM, ASC, CA, ANS
10	AIS, ASM, C.A., ANS
11	AIS, ASM, C.A., ASC, ANS
12	AIS, C.A., ANS
13	AIS, C.A., ASC, ANS
14	AIS, C.A., ASC, ASM, ANS
15	AIS, C.A., ASM, ANS
16	AIS, C.A., ASM, ASC, ANS

Table 4 presents the *CPSImp* values assigned to the sub-systems involved in the discovered attack paths. Note that the *CPSImp* of the AIS, Advanced Sensor Module (ASM), and Collision Avoidance (C.A.) sub-systems is set to *2*, while the *CPSImp* of the Autonomous Navigation System (ANS) and the Autonomous Ship Controller (ASC) sub-systems is set to *3*. This is because the former are navigational systems that provide voyage, dynamic, and static data; the redundancy of such data is sufficient since other on-board systems generate and transmit dynamic and voyage data respectively. Therefore, potential malfunction in any of the AIS, Advanced Sensor Module (ASM), or Collision Avoidance (C.A.) sub-systems cannot cause significant damage to the overall system. On the other hand, the *CPSImp* of the Autonomous Navigation System (ANS) and of the Autonomous Ship Controller (ASC) is *3* since both systems control other navigational systems, and they also have attained the highest *TIC* and *TOC* values, as shown in Table 2.

The application of Steps 5 and 6 of the proposed method on the attack paths of Step 4 yields the prioritized list of attack paths shown in Table 5.

Table 4: Importance of navigational CPSs

CPS	CPSImp
AIS	2
ANS	3
ASM	2
CA	2
ASC	3

Table 5: Prioritized list of attack paths

Path ID	Affected CPSs	Attack Path Importance
6	AIS, ASC, C.A., ASM, ANS	**8.08**
9	AIS, ASM, ASC, CA, ANS	**8.08**
11	AIS, ASM, C.A., ASC, ANS	**8.08**
14	AIS, C.A., ASC, ASM, ANS	**8.08**
4	AIS, ASC, ASM, C.A., ANS	**8.08**
16	AIS, C.A., ASM, ASC, ANS	8.08
5	AIS, ASC, C.A., ANS	6.88
8	AIS, ASM, ASC, ANS	6.88
13	AIS, C.A., ASC, ANS	6.88
3	AIS, ASC, ASM, ANS	6.88
10	AIS, ASM, C.A., ANS	6.28
15	AIS, C.A., ASM, ANS	6.28
2	AIS, ASC, ANS	5.68
7	AIS, ASM, ANS	5.08
12	AIS, C.A., ANS	5.08
1	AIS, ANS	3.88

5 Conclusions

In this work we proposed a novel systematic method for analyzing attack paths in interconnected CPSs. Contrary to existing alternatives, the method handles the scalability problem of attack graphs by considering highly critical nodes and analyzes the resulting paths by considering the cyber risk that each of these represents to the overall system rather than only considering vulnerabilities. We illustrated the workings of the method by applying it to the navigational CPSs of the C-ES, to analyze the possible attack paths that start at the AIS and target the ANS. Five highly critical attack paths have been identified. The results of this analysis can then be fed back to the risk-based process of identifying appropriate countermeasures to satisfy the relevant security requirements and check whether indeed the selected countermeasures alter the possible attack paths and decrease the risk. One pathway for future work is to apply the method as part of an holistic process to identify and analyze cyberattack paths for all the on-board CPSs of the C-ES, so as to propose a complete system security architecture for the C-ES.

References

1. M. N. Al-Mhiqani, R. Ahmad, W. Yassin, A. Hassan, Z. Z. Abidin, N. S. Ali, and K. H. Abdulkareem. Cyber-security incidents: a review cases in cyber-physical systems. *International Journal of Advanced Computer Science and Applications*, 9(1):499–508, 2018.
2. G. Kavallieratos, S. Katsikas, and V. Gkioulos. Modelling shipping 4.0: A reference architecture for the cyber-enabled ship. In *Proceedings of the Asian Conference on Intelligent Information and Database Systems*, pages 202–217. Springer, 2020.
3. G. R. Emad, M. Khabir, and M. Shahbakhsh. Shipping 4.0 and training seafarers for the future autonomous and unmanned ships. In *Proceedings of the 21th Marine Industries Conference (MIC2019)*, pages 202–217, 2020.
4. CH Chang, S Wenming, Z Wei, P Changki, and CA Kontovas. Evaluating cyber-security risks in the maritime industry: a literature review. In *Proceedings of the International Association of Maritime Universities (IAMU) Conference*, 2019.
5. D. M. Silgado. Cyber-attacks: a digital threat reality affecting the maritime industry, 2018.
6. V. Hassani, N. Crasta, and A. M Pascoal. Cyber security issues in navigation systems of marine vessels from a control perspective. In *Proceedings of the ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering*. American Society of Mechanical Engineers Digital Collection, 2017.
7. G. Kavallieratos, S. Katsikas, and V. Gkioulos. Cyber-attacks against the autonomous ship. In *Proceeding of the SECPRE 2018, CyberICPS 2018. Lecture Notes in Computer Science, vol 11387*, pages 20–36. Springer, 2018.
8. Y. C. Chen, V. Mooney, and S. Grijalva. A survey of attack models for cyber-physical security assessment in electricity grid. In *Proceedings of the 2019 IFIP/IEEE 27th International Conference on Very Large Scale Integration (VLSI-SoC)*, pages 242–243. IEEE, 2019.
9. H. S. Lallie, K. Debattista, and J. Bal. A review of attack graph and attack tree visual syntax in cyber security. *Computer Science Review*, 35:100219, 2020.
10. H. Al-Mohannadi, Q. Mirza, A. Namanya, I. Awan, A. Cullen, and J. Disso. Cyber-attack modeling analysis techniques: An overview. In *Proceedings of the 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pages 69–76. IEEE, 2016.
11. Jin B. Hong and Dong Seong Kim. Performance analysis of scalable attack representation models. In *Proceedings of the Security and Privacy Protection in Information Processing Systems*, pages 330–343, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
12. A. Xie, Z. Cai, C. Tang, J. Hu, and Z. Chen. Evaluating network security with two-layer attack graphs. In *Proceedings of the 2009 Annual Computer Security Applications Conference*, pages 127–136, 2009.
13. X. Ou, W. F. Boyer, and M. A. McQueen. A scalable approach to attack graph generation. In *Proceedings of the 13th ACM Conference on Computer and Communications Security*, CCS '06, page 336–345, New York, NY, USA, 2006. Association for Computing Machinery.
14. O. Sheyner, J. Haines, S. Jha, R. Lippmann, and J. M. Wing. Automated generation and analysis of attack graphs. In *Proceedings 2002 IEEE Symposium on Security and Privacy*, pages 273–284, 2002.
15. M. Dacier, Y. Deswarte, and M. Kaâniche. Models and tools for quantitative assessment of operational security. In *Proceedings of the Information Systems*

Security: Facing the information society of the 21st century, pages 177–186, Boston, MA, 1996. Springer US.

16. M. Dacier. *Towards Quantitative Evaluation of Computer Security*. PhD thesis, PhD thesis, Institut National Polytechnique de Toulouse, 1994.
17. M. Dacier and Yves Deswarte. Privilege graph: An extension to the typed access matrix model. In *Proceedings of the Computer Security — ESORICS 94*, pages 319–334, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg.
18. C. Phillips and L. P. Swiler. A graph-based system for network-vulnerability analysis. In *Proceedings of the 1998 Workshop on New Security Paradigms*, NSPW '98, page 71–79, New York, NY, USA, 1998. Association for Computing Machinery.
19. S. Khaitan and S. Raheja. Finding optimal attack path using attack graphs: a survey. *International Journal of Soft Computing and Engineering*, 1(3):2231–2307, 2011.
20. X. Ou and A. Singhal. Attack graph techniques. In *Quantitative Security Risk Assessment of Enterprise Networks*, pages 5–8, New York, NY, 2011. Springer New York.
21. L. P. Swiler, C. Phillips, D. Ellis, and S. Chakerian. Computer-attack graph generation tool. In *Proceedings DARPA Information Survivability Conference and Exposition II. DISCEX'01*, volume 2, pages 307–321 vol.2, 2001.
22. X. Ou, S. Govindavajhala, and A. Appel. Mulval: A logic-based network security analyzer. In *Proceedings of the USENIX security symposium, 2005*, pages 8–8, 07 2005.
23. S. Jajodia, S. Noel, and B. O'Berry. *Topological Analysis of Network Attack Vulnerability*, pages 247–266. Springer US, Boston, MA, 2005.
24. P. Ammann, D. Wijesekera, and S. Kaushik. Scalable, graph-based network vulnerability analysis. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, CCS '02, page 217–224, New York, NY, USA, 2002. Association for Computing Machinery.
25. J. Zeng, S. Wu, Y. Chen, R. Zeng, and C. Wu. Survey of attack graph analysis methods from the perspective of data and knowledge processing. *Security and Communication Networks*, 2019.
26. L. H. Hsu and C. K. Lin. *Graph Theory and Interconnection Networks*. CRC Press, 2019.
27. K. Ingols, R. Lippmann, and K. Piwowarski. Practical attack graph generation for network defense. In *Proceedings of the 22nd Annual Computer Security Applications Conference (ACSAC'06)*, pages 121–130, 2006.
28. K. Kaynar and F. Sivrikaya. Distributed attack graph generation. *IEEE Transactions on Dependable and Secure Computing*, 13(5):519–532, 2016.
29. K. Bi, D. Han, and Jun W. K maximum probability attack paths dynamic generation algorithm. *Computer Science and Information Systems*, 13(2):677 – 689, 2016.
30. N. Poolsappasit, R. Dewri, and I. Ray. Dynamic security risk management using bayesian attack graphs. *IEEE Transactions on Dependable and Secure Computing*, 9(1):61–74, 2012.
31. Jehyun L., Heejo L., and H. P. In. Scalable attack graph for risk assessment. In *Proceedings of the International Conference on Information Networking*, pages 1–5, 2009.
32. F. Dai, Y. Hu, K. Zheng, and B. Wu. Exploring risk flow attack graph for security risk assessment. *IET Information Security*, 9(6):344–353, 2015.

33. John H Castellanos, Martín Ochoa, and Jianying Zhou. Finding dependencies between cyber-physical domains for security testing of industrial control systems. In *Proceedings of the 34th Annual Computer Security Applications Conference*, pages 582–594, 2018.
34. N. Polatidis, M. Pavlidis, and H. Mouratidis. Cyber-attack path discovery in a dynamic supply chain maritime risk management system. *Computer Standards & Interfaces*, 56:74–82, 2018.
35. H. Mouratidis and V. Diamantopoulou. A security analysis method for industrial internet of things. *IEEE Transactions on Industrial Informatics*, 14(9):4093–4100, 2018.
36. M. Ibrahim, Q. Al-Hindawi, R. Elhafiz, A. Alsheikh, and O. Alquq. Attack graph implementation and visualization for cyber physical systems. *Processes*, 8(1):12, 2020.
37. V. Bolbot, G. Theotokatos, E. Boulougouris, and D. Vassalos. Safety related cyber-attacks identification and assessment for autonomous inland ships. In *Proceedings of the International Seminar on Safety and Security of Autonomous Vessels (ISSAV)*, 2019.
38. A. Akbarzadeh and S. Katsikas. Identifying critical components in large scale cyber physical systems. In *Proceedings of the 1st International Workshop on Engineering and Cybersecurity of Critical Systems (EnCyCriS)*, 2020.
39. A. Shostack. *Threat modeling: Designing for security.* John Wiley & Sons, 2014.
40. Microsoft. Chapter 3 – threat modeling. https://docs.microsoft.com/en-us/previous-versions/msp-n-p/ff648644(v=pandp.10)?redirectedfrom=MSDN, note = Accessed: 2020-05-26, 2010.
41. S. D. Zinsmaier, H. Langweg, and M. Waldvogel. A practical approach to stakeholder-driven determination of security requirements based on the gdpr and common criteria. In *Proceedings of the International Conference on Information Systems Security and Privacy ICISSP*, pages 473–480, 2020.
42. CASOS. http://www.casos.cs.cmu.edu/index.php. Accessed: 2019-09-10.
43. Guide for conducting risk assessments. NIST SP 800-30 Rev.1, National Institute of Standards and Technology, Gaithersburg MD, USA, 2012.
44. G. Kavallieratos, V. Diamantopoulou, and S. K. Katsikas. Shipping 4.0: Security requirements for the cyber-enabled ship. *IEEE Transactions on Industrial Informatics*, 16(10):6617–6625, 2020.

11 Article VII: Towards a cyber-physical range [7]

Towards a cyber-physical range

Georgios Kavallieratos
Department of Information Security and Communication Technology, Norwegian University of Science and Technology
Gjøvik, Norway
georgios.kavallieratos@ntnu.no

Sokratis K. Katsikas
Department of Information Security and Communication Technology, Norwegian University of Science and Technology
Gjøvik, Norway
sokratis.katsikas@ntnu.no
&
Open University of Cyprus
Latsia, Nicosia, Cyprus
sokratis.katsikas@ouc.ac.cy

Vasileios Gkioulos
Department of Information Security and Communication Technology, Norwegian University of Science and Technology
Gjøvik, Norway
vasileios.gkioulos@ntnu.no

ABSTRACT

Cyber-physical systems are being increasingly employed in everyday applications, including critical ones. This integration of operational technology systems, originally designed to operate in physical isolation -hence with no or little cyber security defences- with information technology systems, by default meant to be networked, dramatically increases the cyber-attack surface of the resulting composite systems. Thus, the assessment of the security posture of cyber-physical systems, as well as the evaluation of the effectiveness and efficiency of the defensive mechanisms become of paramount importance. Unfortunately, testing cyber security in live real-world cyber-physical systems is not advisable, even when it is possible; hence, the use of testbeds is a necessary alternative. This work surveys cyber-physical testbeds in five major application domains, with an eye towards identifying key features to be subsequently used as input to the process of defining requirements for future cyber-physical testbeds with cyber security posture assessment capability. We then propose a reference architecture for the next generation of cyber ranges, namely the cyber-physical ranges.

KEYWORDS

Cyber-Physical Systems, testbed, Critical Infrastructures, cyber security

ACM Reference Format:
Georgios Kavallieratos, Sokratis K. Katsikas, and Vasileios Gkioulos. 2019. Towards a cyber-physical range. In *5th ACM Cyber-Physical System Security Workshop (CPSS '19), July 8, 2019, Auckland, New Zealand.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3327961.3329532

CPSS '19, July 8, 2019, Auckland, New Zealand

ACM ISBN 978-1-4503-6787-5/19/07…$15.00
https://doi.org/10.1145/3327961.3329532

1 INTRODUCTION

Cyber-physical systems (CPS) are smart systems that include engineered interacting networks of physical and computational components [43]. Examples of CPS include the smart grid, autonomous automobile systems, medical devices, process control systems, automatic pilot avionics [31]. Cyber security attacks targeting CPS come in a significant variety and exhibit a steady -and alarming- growth [20]. The impact of such attacks is significant, due to the mission-critical nature of the affected CPS applications [3].

Thus, the study of potential threats and vulnerabilities, and the assessment of the risk that potential cyber-attacks against cyber-physical systems pose, as well as the effectiveness of protection measures, is of paramount importance. Unfortunately, in vivo experimentation with the cyber security of real-world CPS is neither advisable nor even possible.

Therefore, a need for using testbeds for conducting experiments for vulnerability analysis, for testing defense mechanisms, for impact assessment, for threat analysis, and for cyber security tests in general arises [24], [25]. Such testbeds may also be used for training personnel involved with the operation of CPSs in cyber security. CPS testbeds in domains such as smart grids, water distribution, vehicular and transportation systems and medical devices have been reported in the literature. These testbeds are virtual, physical or hybrid, they have been built with different objectives, and have varying capabilities.

Even though a number of testbeds have been developed for studying different cyber-physical systems, only few have been designed to allow cyber security experimentation; allowing for such functionality is an architectural design issue. According to [44], "Cyber ranges are interactive, simulated representations of an organization's local network, system, tools, and applications that are connected to a simulated Internet level environment. They provide a safe, legal environment to gain hands-on cyber skills and a secure environment for product development and security posture testing." Accordingly, a testbed with functionality allowing the testing of the security posture of cyber-physical systems would constitute a "cyber-physical range".

In this paper we propose a reference architecture for a cyber-physical range. A reference architecture is an authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions

[46]. We first survey existing CPS testbeds with cyber security testing capacity, we identify common architectural features, and we define requirements enabling the assessment of cyber-security posture[1] in such testbeds.

We focus on testbeds for the smart grid, industrial control systems, the Internet of Things, vehicular and transportation systems, and medical devices. Several testbeds of SCADA systems have been also identified in the literature [2, 10, 40, 42, 50, 60], as well as a survey of such testbeds [49]. However, SCADA systems constitute only elements of industrial control systems; accordingly testbeds of standalone SCADA systems are outside the scope of this work.

The contribution of this paper is as follows:

- An up-to-date review of existing testbeds for cyber-physical systems with cyber security testing capability, and an analysis of their architecture and key features;
- Requirements of a cyber-physical range; and
- A reference architecture for a cyber-physical range.

The remainder of this paper is structured as follows: Section 2 presents the related work. In section 3 we review existing cyber-physical testbeds with cyber-security testing capability and we discuss important features of these testbeds. In Section 4 we define requirements of a cyber-physical range and we propose an architectural reference model for it. Finally, section 5 summarizes our conclusions.

2 RELATED WORK

We identified research papers and reports for cyber-physical system testbeds by searching in the ACM Digital Library, Science Direct, Scopus, IEEE Xplore and Semantic Scholar databases with appropriate keywords. For the selection of articles we applied the following criteria:

- The article must be exclusively related to cyber-physical system testbeds.
- The article must be directly related to cyber-security testing.

The initial search identified 86 articles. By applying the aforementioned criteria, our analysis focused on thirty two testbeds.

Several testbeds have been developed in order to virtually or physically simulate or emulate cyber physical systems and some relevant surveys have appeared in the literature. M. H. Cintuglu et al. in [11] have surveyed cyber-physical smart grid testbeds and have provided a taxonomy and insightful guidelines for the development of such testbeds. This survey has categorized the testbeds according to their domains, research goals, test platforms and communication infrastructures. H. Holm at al. in [25] have conducted a survey of industrial control system testbeds. Their work reviews 30 testbeds, most of which aim to identify vulnerabilities, educate operators and test different defense mechanisms. In addition, they have categorized the testbeds considering their objectives and fidelity. However, none of these works has considered the ability of testbeds to be used for assessing the security posture of cyber-physical systems. Siatterlis and Genge in [53] present a non-exhaustive review of some CPS testbeds, used in critical infrastructures, and describe the Experimentation Platform for Internet Contingencies (EPIC) testbed; this can be considered as one early instance of a cyber-physical range. On the other hand, the Australian Department of Defense in [13] published an extensive survey of (conventional, non-CPS) testbeds up to 2013. This was complemented by [61], where the relevant literature up to 2017 was reviewed and the description of the KYPO cyber range was provided. Another such survey is provided in [48].

To the best of our knowledge, no published work makes reference to a "cyber-physical range".

[1] The cyber security posture is the status of a critical infrastructure's networks, information, and systems based on information assurance resources and capabilities in place to manage cyber security incidents [45].

3 CYBER - PHYSICAL TESTBEDS WITH CYBER SECURITY TESTING CAPABILITY

In this section CPS testbeds are discussed, categorized according to their application domain. In each category, the reviewed testbeds are sorted in chronological order.

3.1 Smart Grid Testbeds

A smart grid is an upgraded electricity network depending on two-way digital communications between supplier and consumer that in turn give support to intelligent metering and monitoring systems. Smart grids give clear advantages and benefits to the whole society, but the dependency on computer networks and the Internet into future grids makes them more vulnerable to malicious attacks with potentially devastating results [16].

I.N. Fovino et al. in [17] describe a testbed deployed in Leghorn (Livorno, Italy), for identifying vulnerabilities and threats; for developing attack scenarios; for studying the effects of Information and Communication Technology (ICT) attacks against SCADA systems; and for testing cyber security countermeasures in a typical power plant. Its architecture has six main functional elements, namely (i) Field network; (ii) Process network; (iii) Demilitarized Zone; (iv) External network; (v) Observer System; and (vi) Horizontal Services.

The testbed described by *R. Liu et al.* in [36] was designed to identify potential vulnerabilities; simulate real world scenarios and estimate their risks; test the adoption of new applications and hardware devices; educate stakeholders; and analyze the impact of three possible cyber-attacks on a power grid. Architecturally, it comprises four distinct layers, namely (i) the Power system layer, which comprises all simulated power systems; (ii) the Sensor layer which consists of Phasor Measurement Units (PMUs) and Phasor Data Concentrator (PDC); (iii) the Communication layer, where all communication technologies are included; and (iv) the Application layer, which includes an energy management system (EMS).

A real-time testbed environment to validate security techniques in power grids has been proposed by *G. Koutsandria et al.* in [34]. The purpose of this testbed was to validate a proposed network intrusion detection system for industrial control processes. The testbed has been developed using an integrated framework of tools for process management and cyber-physical security, and consists of both physical (real PLCs which communicate via Ethernet), and virtual (simulated) components.

J. Hong et al. in [26] introduce a cyber-physical testbed to study the cyber-physical security of power systems. This testbed aims to study and understand cause-effects relationships of cyber-attacks, vulnerabilities and resilience of power systems and the performance

and reliability of applications. Its architecture has three main modules, namely (i) Physical system module; (ii) ICT module; and (iii) Cyber System module. In more detail, the physical module consists of four types of Intelligent Electronic Devices (IED); the ICT module implements industrial communication protocols (such as ICCP, DNPi, OPC, and IEC 61850); and the cyber module comprises network devices, computer servers, databases, user interfaces, Operator Training Simulator, and Energy Management System.

A. Ashok at al. in [7] describe a remotely accessible testbed for the PowerCyber CPS testbed which is located at Iowa State University. The power simulations are performed via Real Time Digital Simulator (RTDS), Opal-RT and DIgSILIENT Power factory software, and visualization technologies have been employed for scalability. This testbed was developed with an eye towards identifying potential vulnerabilities in power grids; simulating cyber-attack scenarios using ISEAGE; determining the impact of potential cyber-attacks; and evaluating studies using RTDS. The testbed architecture comprises three layers, namely (i) Information, (ii) Communication and (iii) Physical.

E. Tebekaemi et al. in [58] propose a simulation testbed which supports extensive analysis of communication protocols; cyber-physical security functions; ICT vulnerabilities; network configurations; and physical security requirements for an IEC61850-based power distribution substation, by means of developing cyber-attacks and studying their effects on the power substation network. The testbed consists of both virtual and physical components.

In [23], *P. Gunathilaka et al.* describe SoftGrid, a software-based Smart Grid testbed to evaluate the effectiveness, performance, and interoperability of various security solutions which aim to secure the remote control interface of Smart Grid's substations. By conducting experiments in this virtual testbed, stakeholders are able to perform cyber-attacks, assess their implementation, and educate the operators. Furthermore, cyber security solutions such as IDSs, Firewalls, and Security-enhanced Gateways can be validated by using this testbed.

S. Poudel et al. in [47] proposed a real-time cyber-physical system testbed designed to assess cyber security and control stability. The main goals of this testbed are to provide a platform for vulnerability assessment; development of disturbance scenarios; analyze the impact of such attacks; study the stability and the control of the power grid; assess cyber-physical metrics; examine defense mechanisms; and train and educate stakeholders. This testbed consists of both cyber and physical components and uses the SEL-C662 and DNP3.0 communication protocols.

The FUSE testbed for conducting security experiments on smart grids was proposed by *E. Xypolytou et al.* in [62]. FUSE aims to identify methods for micro grid self-management and methods to autonomously control the interactions in a grid, focusing on the security and reliability of communications. It also enables the evaluation of defense mechanisms to prevent and mitigate cyber-attacks. The testbed uses PMUs in order to synchronize and measure frequency, voltage and current phasors. IP connections among the components are established using the Ethernet protocol.

A. Siddiqi et al. in [55] give an overview of the network services provided by industrial devices found in the EPIC testbed at Singapore University of Technology and Design (SUTD). EPIC is an Electric Power ICS system testbed that aims to provide useful information on system vulnerabilities, potential cyber-attacks, attack vectors and countermeasures. The EPIC testbed allows the analysis of several physical processes, such as power production, transmission, and distribution.

3.2 Industrial Control Systems Testbeds

Industrial Control Systems (ICS) are control systems used to facilitate the control, monitoring and production of industrial processes. A key element of an ICS is a SCADA (Supervisory Control and Data Acquisition) system, that aims to continuously monitor and control the underlying controlled process [16].

Sandia National laboratory developed a Virtual Control System Environment (VCSE) to analyze a control system's functionality and operations; explore system vulnerabilities; develop testing protection and mitigation techniques; and understand possible impacts of particular cyber threats [37]. The testbed comprises Virtual Machines (VMs) which host real physical components aiming to simulate models of critical infrastructures, and emulate different interfaces with SCADA systems. Further, by leveraging the OPNET software, the testbed is able to model complex critical infrastructure networks, to identify and analyze potential vulnerabilities. It is worth noting that the specific testbed takes into consideration the human element as a key part for the cyber defense and hence provides an appropriate training environment for building people's skills. This functionality is achieved by using the Umbra tool to simulate complex physical and human-involved systems.

T. Morris et al. in [41] describe the Mississippi State University's SCADA Security Laboratory and Power and Energy Research testbed. The aim of this testbed is to identify vulnerabilities in, perform cyber-attacks against, estimate the risk of such attacks, and develop necessary defense mechanisms for industrial control systems. An additional important goal is to be used by the university for student education. The testbed is built with both physical and virtual components and it is able to simulate electricity infrastructures, gas pipelines, factory systems and water storage and distribution systems.

B. Reaves et al. in [51] proposed a virtual testbed for ICS in order to assess intrusion detection systems. The architecture consists of virtual devices (MTUs, RTUs), actual devices (wireless radios, HMI), configuration files (communication protocols) and data loggers. All components are simulated using python. The authors have also evaluated the performance of the testbed by conducting virtual attacks. However, due to the limited installation of field devices, the flexibility of the testbed is also limited.

NIST in [9] described a cybersecurity testbed aiming to estimate the performance of critical infrastructure systems which have already been designed with a security perspective. Another goal of this testbed is to estimate the impact of new security technologies such as fingerprint and radio frequency fingerprinting. The testbed comprises both physical and virtual components, and its architecture is based on the Tennessee Eastman model. However, the implementation status of this testbed is not clear.

E. E. Miciolino et al. in [39] introduce a testbed for monitoring and control of a water system, with an eye towards examining the security of the Modbus/TCP protocol. Its aim is to detect, identify and analyze physical faults, cyber-attacks and anomalies which

may occur in critical infrastructures. By leveraging this testbed, researchers have performed experiments to assess the consequences of a cyber-attack in the SCADA communication network.

I. Ahmed et al. in [2] describe a testbed built at the University of New Orleans. This testbed simulates three industrial physical processes, namely (i) a gas pipeline; (ii) a power transmission and distribution system; and (iii) a wastewater treatment plant. The aim of the testbed is to provide a resource for cybersecurity research, forensic research and education on industrial control systems. For the gas pipeline the testbed consists of one analog pressure gauge; one solenoid valve; a digital pressure Gauge and Transmitter; one manual valve; and one air compressor. These components are connected to a PLC using an Ethernet switch. On the other hand, the power transmission testbed consists of a power station, four substations and five voltmeters. All components are connected to the control center using a PLC and the Modbus protocol. In the wastewater treatment testbed, a sedimentation tank, water level sensor, aeration tanks, and clarification tanks are used. The communication between the control center and the PLC of the wastewater simulator is by means of the PROFINET protocol.

A simulation-based testbed, applicable to evaluating security on different critical infrastructure domains has been proposed by *A. Ghaleb* in [19]. Two realistic settings have been simulated, namely a water distribution grid and an Electrical power grid, using OMNet++ software.

W. Hurst et al. in [28] propose a testbed which brings together physical and virtual tools of a water distribution plant in order to educate students and researchers on potential cyber-attacks that may occur in such infrastructure. This testbed's approach combines virtual and real components in a cost efficient way.

E. Korkmaz et al. describe the Binghamton Testbed in [32], and a delay attack case study on it [33]. This testbed was developed by the Office of Naval Research to provide an infrastructure for experimentation by performing various cyber-attacks and analyzing different mitigation mechanisms in ICS. The important component of this testbed is the tools that contain preconfigured network attack scenarios. The testbed comprises both physical and virtual components.

The SWaT testbed was designed by *A. P. Mathur et al.* [38], to identify potential vulnerabilities and to study cyber and physical attacks and defense mechanisms on a water treatment system. An additional important goal of this testbed is to educate operators on how to handle such incidents and how these attacks cascade between ICS components. The testbed consists of physical components, such as an Ultrafiltration Unit, a Chemical dosing station, a UV dechlorinator, a Reverse Osmosis Units and a cabinet of PLCs. The PLCs communicate with sensors and actuators, among themselves and with SCADA servers and other computers, by using ring based Ethernet topology. Industrial switches, HMI, the SCADA server and the Historian server are connected using star based Ethernet topology.

C. M. Ahmed et al. in [1] propose a water distribution testbed (WADI) for research in the design of secure CPSs. The testbed's goals are to allow conducting security experiments in the infrastructure, evaluating different defense mechanisms, and understanding the impact of cascading attacks. WADI's communication architecture is similar to that of SWaT. Although the two testbeds have similar communication architectures, their physical components are different. Namely, WADI consists of two Water tanks, level sensors, chemical dosing system, sensors, two reservoir tanks, actuators, RTUs and PLCs. The PLCs are programmed by the National Instruments (NI) LabVIEW software, whilst the RTUs are configured using a Schneider Electric SCADA pack workbench. The main communication protocol used is the Modbus.

In [4], *M. Almgren et al.* present the national testbed for security research in Sweden. The testbed has been implemented by Resilient Information and Control Systems (RICS) over the Cyber Range and Training Environment (CRATE) at the Swedish Defence Research Agency (FOI). The entire testbed is virtual, and uses virtualization environments such as VirtualBox. Although the testbed is still under development, it has been used to generate data in order to create traffic patterns for anomaly detection. In addition, the testbed has been used within the iPilot project, which trains Swedish nuclear power plant operators.

B. Green et al. in [22] present an extensive ICS testbed for security research in Lancaster university. The testbed comprises physical components and virtual platforms such as SCADA historians and workstations aiming to examine the resilience of the ICS and utility networks. The combination of physical and virtual components facilitates the integration of new systems towards a more scalable architecture. The testbed architecture is based on the Purdue Enterprise Reference Architecture, and takes into account the development of systems and devices across Levels 1, 2, 3, and 4 of the reference model. The authors also conducted a survey among similar physical testbeds and they noticed the lack of process diversity and of simulation support in these.

3.3 Internet of Things Testbeds

Internet of Things testbeds focus mostly on smart home applications. A smart home environment deploys several devices with diverse functionality, all connected to the network. Such devices could be smartphones, alarm systems, cooling systems and power control and monitor systems.

M. A. Crossman et al. in [5] implemented a testbed to emulate the IoT in order to assess a two-factor authentication mechanism that they proposed.

A. Tekeoglu et al. in [59] propose a testbed to investigate security and privacy issues of IoT devices, based on off-the-self hardware and open source software. Through this implementation, the authors capture transmitted data packets and analyze a wide range of security and privacy issues. The testbed facilitates vulnerability analysis, firmware/application updates and cloud security analysis.

The testbed described in [54] has been designed to allow the analysis of IoT devices, against established security requirements. The testbed follows a layer-based platform model, with modular structure; this allows testing any smart device without the need to make modifications to the infrastructure. There are four main modules in the testbed, namely (i) the management and report module; (ii) the security testing manager module; (iii) the security testing module; and (iv) the measurements and analysis module.

Another testbed for modelling and assessing the security of the IoT has been proposed by *M. Ge et al.* in [18]. Its operation goes through five phases, namely (1) data processing; (2) security model

generation; (3) security visualization; (4) security analysis; and (5) model updates. Through these phases, an operator is able to identify potential attack scenarios, to analyze the security of the IoT, and to assess the effectiveness of different countermeasure selection options.

W. Hurst et al. in [27] introduce a testbed for cyber-security and training in the IoT. The testbed incorporates various components, such as Arduino, client PC, Web server, three IoT devices, and an Arduino WiFi shield; this allows the collection of useful data from different components. Cyber-attack scenarios have been developed by leveraging this testbed.

3.4 Vehicular and transportation systems testbeds

Vehicular and transportation systems include applications such as smart cars, intelligent road systems, smart ships and smart trains.

N. H. Desso [15] introduces the Naval Postgraduate School's Machinery Control Systems (MCS) testbed, designed to provide a working model that replicates the MCS systems in the U.S. Navy fleet. The overall goal of this testbed is to help ship designers in the U.S. Navy to prepare for potential cyber-attacks, as well as to enable researchers to conduct experiments aiming at ensuring the vessel's safe operation. The testbed consists of real MCS components, thus allowing for realistic cyber security testing and research. Specifically, the machinery part contains two test systems: (i) an analog fluid tank lab (AFTL) and (ii) a digital I/O lab (DIOL). These systems are connected to PLCs, the HMI, and the PLC workstation.

X. Zhen et al. in [63] propose a real time simulation environment to integrate the simulated Control Area Network (CAN) bus system with an emulated infotainment system in order to identify potential vulnerabilities in automobiles employing the CAN bus. Electronic Control Units (ECUs) are the main controller units for the vehicle; these use the CAN bus to exchange packets. The Vehicle Control Unit (VCU) is a dynamic model of the simulated electric vehicle. Its aim is to navigate the simulated vehicle according to the current position and speed messages published on the CAN bus by the ECUs. The CAN gateway is responsible for providing two main functions; an in-vehicle CAN to Ethernet gateway and a CAN to an Ethernet data capture gateway. Additionally, an infotainment Unit acts as a conduit for studying remote attacks against the CAN bus.

3.5 Medical devices testbeds

Medical devices include all systems that gather and analyze health data, such as body sensors and implantable devices.

Y. Berhanu et al. in [8] propose a testbed for WBANs (Wireless Body Area Network) in eHealth applications, that uses current COTS products and open source software, and aims to be used to study risk-based adaptive security methods and mechanisms for IoT in Health. The testbed consists of six Shimmer nodes, one TelosB, two Wizzmotes and two Rasberry Pis. Smartphones serve as concentrators or sink nodes, that gather eHealth readings for sensor nodes, whilst computers and tablets are used for backend storage and analysis. The main communication technology is the Ethernet.

3.6 Testbed features, limitations, and challenges

Based on the survey in the previous section, in this section we identify and discuss certain features of the testbeds; these are summarized in Table 1, and will be used as the basis to define requirements and to introduce the reference architecture model of the cyber-physical range in the next section.

Three testbed implementation approaches have been identified, namely (i) Physical, (ii) Virtual and (iii) Hybrid. Physical testbeds are expensive and time-consuming to build, not portable and hard to maintain. Additionally, creating a full-scale physical replica of large scale CPSs is not practical. Hence, it is not surprising that the majority of the surveyed testbeds have opted for software-based solutions, to reduce costs and increase flexibility. However, virtual testbeds cannot support hands-on experience, many physical attacks cannot be implemented, and physical impact cannot be assessed. Therefore, a hybrid testbed, deploying a mix of emulation, simulation, and physical components appears to be the best option. However, there is a high risk that the built testbed does not replicate accurately the CPS [12].

Several security testing functionalities have been identified. Most of the described testbeds can be used for vulnerability analysis. Six of the existing testbeds were developed to assess the impact of potential cyber-attacks by leveraging the testbed to conduct experiments. On the other hand, only two among the surveyed testbeds have threat analysis capability, by allowing the development and testing of different attack scenarios. Only few testbeds incorporate defensive mechanisms, even though the capability to adopt various security mechanisms such as Intrusion Detection Systems exists. All examined testbeds aim to contribute to the security training of the operators. Through these testbeds, operators are able to understand and assess the security posture of the relevant CPSs.

Several limitations of the surveyed testbeds have been identified. The testbeds described in [7, 17, 32, 36] have examined a limited number of attack models and datasets. Many testbeds, such as [7, 17, 47, 55], are not flexible or scalable; thus they cannot support different devices and the services these offer, since they have been designed to reflect a specific instance of a CPS. On the other hand, testbeds such as [28, 33, 34, 39, 62] have focused on the data flows and the communication among the components; hence, they have not examined potential cyber-physical attacks. Further, [58] and [23] employ a static communication architecture; thus they are limited in the variety of communication protocols they can employ. In contrast, such tests can be performed on the virtual testbeds described in [51] and [4]. Several testbeds [1, 2, 38] cannot develop to full-scale, therefore they are incapable of reflecting accurately the entire CPS, viewed as a system-of-systems. Consequently, large-scale attacks which could compromise many devices over the infrastructure hosting the CPSs cannot be examined and analyzed.

4 REFERENCE ARCHITECTURE FOR A CYBER-PHYSICAL RANGE

4.1 Requirements

Vykopal et al [61] identified the *must have* requirements of a cyber range. Building upon these, and considering the nature and the

	Implementation approaches			Security testbed functionality				
	Physical	**Virtual**	**Hybrid**	**Vulnerability Analysis**	**Training**	**Defensive Mechanism**	**Assessment of cyber attack impact**	**Threat Analysis**
[7]			✓	✓	✓		✓	
[36]			✓	✓	✓		✓	
[17]			✓	✓	✓			✓
[34]			✓	✓	✓			
[26]		✓		✓	✓			
[32]			✓	✓	✓			
[55]			✓	✓	✓	✓		✓
[47]			✓		✓	✓	✓	
[23]		✓			✓		✓	
[38]	✓			✓	✓	✓		
[1]	✓			✓	✓	✓	✓	
[62]			✓		✓	✓		
[41]			✓	✓	✓			
[9]			✓		✓		✓	
[51]		✓			✓			
[2]			✓		✓			
[56]		✓			✓			
[39]			✓		✓			
[33]			✓		✓		✓	
[28]		✓			✓			
[4]			✓		✓			
[19]		✓			✓			
[27]			✓		✓			
[18]		✓			✓			
[59]	✓			✓	✓			
[54]			✓	✓	✓			
[52]	✓				✓			
[5]			✓		✓			
[63]		✓		✓	✓			
[14]			✓	✓	✓			
[8]		✓			✓			
[22]			✓		✓		✓	

Table 1: Testbed features

characterics of a cyber-physical range, we define the following requirements for a cyber-physical range:

Flexibility: The cyber-physical range should be able to handle different CPSs, to allow the exchange of information and/or components among these, and to perform various functions of the CPSs it reflects.

Scalability: The cyber-physical range should scale well in terms of the number of CPS components (devices and services), processing power and other available resources of the individual components, and the number of users.

Isolation: The cyber-physical range and its users should be isolated from the outside world and from each other.

Interoperability: The cyber-physical range should be able to connect to, integrate with, and to work cooperatively with external systems, with reasonable effort.

Cost-Effectiveness: The cyber-physical range should support deployment of hardware while keeping the operational and maintenance costs as low as possible. Open source software should be used as much as possible.

Built-In Monitoring: The cyber-physical range should natively provide both real-time and post-mortem access to detailed monitoring data. These data should include flow data and captured packets from the network links, as well as metrics and logs related to the process that the CPS reflects.

Easy Access: Experienced users should be able to use the cyber-physical range with reasonable training.

Adaptability: Due to the wide range of cyber-physical systems that the cyber-physical range should be able to reflect, it should be possible to install and uninstall different components in the testbed with reasonable re-configuration effort.

Shareability: It should be possible for individual components of the cyber-physical range to be shared.

4.2 Reference architecture

Building upon the analysis of the surveyed testbeds and on the identified requirements, Figure 1 depicts a proposal for a reference architecture of a cyber-physical range. The architecture comprises four main modules: (i) The Testbed control center, (ii) the physical components, (iii) the virtual components, and (iv) the cybersecurity defensive mechanisms. These are discussed in the sequel.

- **Testbed control center module:** This module describes the interaction between the operator and the range.
- **Physical components module:** This includes the range's physical components. Examples of such components are PLCs, IEDs, RTUs, smartphones, tablets, smartwatches, IP cameras, ECUs, vehicle telemetry displays, Raspberry PIs, medical devices, etc. The module also contains sensors and actuators, as well as switches and routers. Various communication protocols should be possible to use for communication among the components.
- **Virtual components module:** This module contains the emulated/simulated components necessary for emulating/simulating CPSs in different operational domains. Figure 2 depicts four such components, one each for a power grid/ICS environment, an IoT environment, a vehicular system environment, and a medical devices environment.
- **Cybersecurity defensive mechanisms module:** This module includes a collection of defensive mechanisms to be used for assessing the cyber-security of the CPSs of interest. Examples of such mechanisms are Intrusion Detection Systems (e.g. as proposed in [35]), Firewalls (e.g. as described in [6, 23]), Security enhanced gateways (e.g. as in [23]), other configurable security devices (such as e.g. crypto devices [29]) etc.

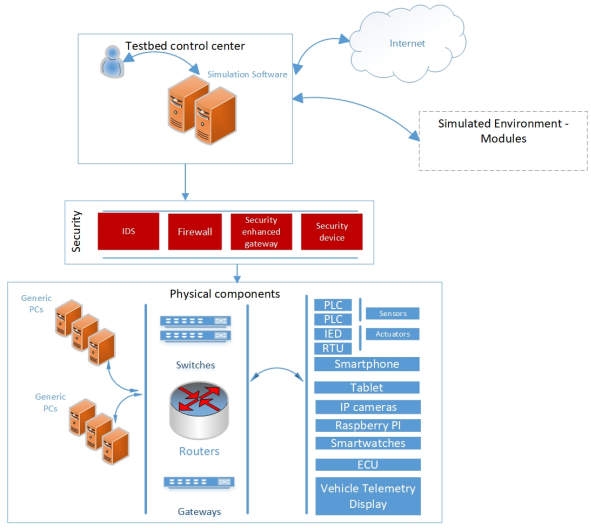


Figure 1: Reference architecture

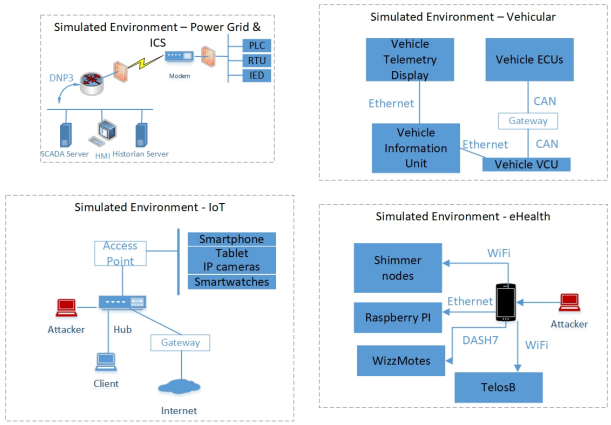


Figure 2: Simulated environments

An example of the instantiation of the proposed reference arhitecture is a testbed for testing the cyber security posture of the navigation systems of the Cyber-enabled ship (C-ES), as they have been described in [30]. To facilitate the analysis of the vessel's navigation systems, the testbed must provide data and information from its environment, the maritime traffic in the vicinity, and the control commands such as speed and rudder directions. To this end, a set of physical and virtual components can be used to create an appropriate testbed, such as the one shown in Figure 3. Namely, an Automatic Identification System (AIS), an Electronic Chart Display and Information System (ECDIS), a RADAR, and a set of routers and switches constitute the physical components of the testbed, since these have been found to be the most vulnerable systems [30, 57]. By leveraging a simulation software, a set of data structures can be developed to simulate the functions of the collision avoidance system and the operations of the Integrated Bridge System (IBS), and hence identify potential risks which derive from data exchanges among the systems. Moreover, an offshore bridge simulator can be used to extract the necessary data from the ship's environment. Such data can, for example, be derived from engine control and

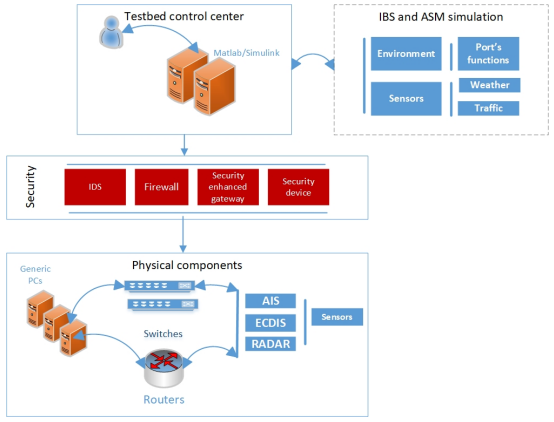


Figure 3: C-ES's Navigation Testbed

monitoring systems and from several sensors in the vessel that provide weather and traffic data.

A crucial part of the security analysis is the examination of the information exchange among systems; this is based on the S-100 model proposed by DNV GL in [21]. Additionally, a Security Information and Event Management system (SIEM) and an IDS can be used for monitoring and assessing the functionality of both the physical and virtual components of the testbed.

Such a testbed allows to launch cyber-attacks against both physical (AIS) and virtual (IBS) components, to analyze Human Machine Interfaces (HMI), and to assess systems interconnections, dependencies and interdependencies. Thus, the testbed provides a platform that may be used to validate different attack scenarios and understand the propagation of the risks within the ship infrastructure, by using multiple data sources for route planning monitoring, and navigation. For instance, the integrity and the confidentiality of different system functions, such as the update/amendment of the current route shown on the ECDIS monitor may be evaluated. Furthermore, the experimentation with different security mechanisms such as firewalls and IDS is enabled, so as to enable the assessment of the efficiency of the approach for mitigating the identified risks. The testbed can also be used to increase the awareness of the operators and stakeholders of potentially complicated environmental conditions and threat scenarios, by simulating navigation operations and analyzing the security of the bridge network and of the Integrated Bridge System (IBS). Additionally, the examination of potential security implications during system maintenance or replacement operations is made possible.

5 CONCLUSIONS

We surveyed existing testbeds of cyber-physical systems with cyber security posture assessment capability in five major application domains. Despite their differences, that stem form the diverse application domains, certain common key features have been identified. we used these as the basis for defining requirements that a generic cyber-physical testbed with security posture assessment capability should satisfy, and we coined the term "cyber-physical range" for such a testbed. We then defined a reference model architecture for such a range, which can be instantiated to any of the surveyed testbeds. This model comprises four main modules, and incorporates both physical and virtual devices, as well as cyber defensive mechanisms, thus allowing the creation of a hybrid cyber-physical testbed with cyber security posture assessment capability. An example of the instantiation of the proposed reference architecture to the case of the navigation systems of a cyber-enabled ship has been provided. Our future work plans include the refinement of the testbed reference model architecture; its instantiation to specific domain environments, including one for an remotely controlled vessel and the associated shore control center; the design and development of a modular cyber-physical range, its use for experimentation and validation of its effectiveness, efficiency, configurability, and performance.

REFERENCES

[1] Chuadhry Mujeeb Ahmed, Venkata Reddy Palleti, and Aditya P Mathur. 2017. WADI: A water distribution testbed for research in the design of secure cyber physical systems. In *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*. ACM, 25–28.

[2] Irfan Ahmed, Vassil Roussev, William Johnson, Saranyan Senthivel, and Sneha Sudhakaran. 2016. A SCADA system testbed for cybersecurity and forensic research and pedagogy. In *Proceedings of the 2nd Annual Industrial Control System Security Workshop*. ACM, 1–9.

[3] Al Balushi T. Nadir Z. Hussain O.K. Ali, S. 2018. *Cyber Security for Cyber Physical Systems.*

[4] Magnus Almgren, Peter Andersson, Gunnar Björkman, Mathias Ekstedt, Jonas Hallberg, Simin Nadjm-Tehrani, and Erik Westring. 2018. RICS-el: Building a National Testbed for Research and Training on SCADA Security (Short Paper). In *International Conference on Critical Information Infrastructures Security*. Springer, 219–225.

[5] M. A. Crossman and. 2015. Study of authentication with IoT testbed. In *2015 IEEE International Symposium on Technologies for Homeland Security (HST)*. 1–7. https://doi.org/10.1109/THS.2015.7225303

[6] Tofino Security Appliance. [n. d.]. Protect your SCADA and Industrial Control Systems Against Network Problems and Cyber Threats.

[7] Aditya Ashok, Sujatha Krishnaswamy, and Manimaran Govindarasu. 2016. PowerCyber: A remotely accessible testbed for Cyber Physical security of the Smart Grid. In *Innovative Smart Grid Technologies Conference (ISGT), 2016 IEEE Power &*

Energy Society. IEEE, 1–5.

[8] Yared Berhanu, Habtamu Abie, and Mohamed Hamdi. 2013. A Testbed for Adaptive Security for IoT in eHealth. In *Proceedings of the International Workshop on Adaptive Security (ASPI '13)*. ACM, New York, NY, USA, Article 5, 8 pages. https://doi.org/10.1145/2523501.2523506

[9] Richard Candell, Keith Stouffer, and Dhananjay Anand. 2014. A cybersecurity testbed for industrial control systems. In *Proceedings of the 2014 Process Control and Safety Symposium*.

[10] Wang Chunlei, Fang Lan, and Dai Yiqi. 2010. A Simulation Environment for SCADA Security Analysis and Assessment. 342 – 347. https://doi.org/10.1109/ICMTMA.2010.603

[11] Mehmet Hazar Cintuglu, Osama A Mohammed, Kemal Akkaya, and A Selcuk Uluagac. 2017. A Survey on Smart Grid Cyber-Physical System Testbeds. *IEEE Communications Surveys and Tutorials* 19, 1 (2017), 446–464.

[12] Edward JM Colbert and Alexander Kott. 2016. *Cyber-security of SCADA and other industrial control systems*. Vol. 66. Springer.

[13] Jon Davis and Shane Magrath. 2013. *A survey of cyber ranges and testbeds*. Technical Report. DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION EDINBURGH (AUSTRALIA).

[14] A. Dembovskis. 2012. Testbed for performance evaluation of SAT-AIS receivers. In *2012 6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC)*. 253–257. https://doi.org/10.1109/ASMS-SPSC.2012.6333085

[15] N. H. Desso. 2014. Designing a Machinery Control System (MCS) Security testbed, Thesis.

[16] ENISA. [n. d.]. Critical Infrastructures and Services.

[17] Igor Nai Fovino, Marcelo Masera, Luca Guidi, and Giorgio Carpi. 2010. An experimental platform for assessing SCADA vulnerabilities and countermeasures in power plants. In *Human System Interactions (HSI), 2010 3rd Conference on*. IEEE, 679–686.

[18] Mengmeng Ge, Jin B. Hong, Walter Guttmann, and Dong Seong Kim. 2017. A framework for automating security analysis of the internet of things. *Journal of Network and Computer Applications* 83 (2017), 12 – 27. https://doi.org/10.1016/j.jnca.2017.01.033

[19] A. Ghaleb, S. Zhioua, and A. Almulhem. 2016. SCADA-SST: a SCADA security testbed. In *2016 World Congress on Industrial Control Systems Security (WCICSS)*. 1–6. https://doi.org/10.1109/WCICSS.2016.7882610

[20] Jairo Giraldo, Esha Sarkar, Alvaro A Cardenas, Michail Maniatakos, and Murat Kantarcioglu. 2017. Security and privacy in cyber-physical systems: A survey of surveys. *IEEE Design & Test* 34, 4 (2017), 7–17.

[21] DNV GL. 2015. Ship connectivity. https://doi.org/Positionpaper

[22] Benjamin Green, Anhtuan Lee, Rob Antrobus, Utz Roedig, David Hutchison, and Awais Rashid. 2017. Pains, gains and PLCs: ten lessons from building an industrial control systems testbed for security research. In *10th {USENIX} Workshop on Cyber Security Experimentation and Test ({CSET} 17)*.

[23] Prageeth Gunathilaka, Daisuke Mashima, and Binbin Chen. 2016. Softgrid: A software-based smart grid testbed for evaluating substation cybersecurity solutions. In *Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy*. ACM, 113–124.

[24] Wes Hardaker, Darrell Kindred, Ron Ostrenga, Dan Sterne, and Roshan Thomas. 2002. Justification and requirements for a national DDoS defense technology evaluation facility. *Network Associates Laboratories Report* (2002), 02–052.

[25] Hannes Holm, Martin Karresand, Arne Vidström, and Erik Westring. 2015. A survey of industrial control system testbeds. In *Secure IT Systems*. Springer, 11–26.

[26] Junho Hong, Ying Chen, Chen-Ching Liu, and Manimaran Govindarasu. 2015. Cyber-physical security testbed for substations in a power grid. In *Cyber Physical Systems Approach to Smart Electric Power Grid*. Springer, 261–301.

[27] William Hurst, Nathan Shone, Abdennour El Rhalibi, Andreas Happe, Ben Kotze, and Bob Duncan. 2017. Advancing the Micro-CI Testbed for IoT Cyber-Security Research and Education. In *Eighth International Conference on Cloud Computing, GRIDs, and Virtualization*, Carlos Becker Westphall, Yong Woo Lee, Bob Duncan, Aspen Olmsted, Michael Vassilakopoulos, Costas Lambrinoudakis, Sokratis K. Katsikas, and Raimund Ege (Eds.). IARIA, 129–134.

[28] William Hurst, Nathan Shone, Qi Shi, and Behnam Bazli. 2016. MICRO-CI: A Testbed for Cyber-Security Research. In *EMERGING 2016: The Eighth International Conference on Emerging Networks and Systems Intelligence*. IARIA XPS Press, 17–22.

[29] Sungmo Jung, Jae-Gu Song, and Seoksoo Kim. 2008. Design on SCADA testbed and security device. *International Journal of Multimedia and Ubiquitous Engineering* 3 (11 2008).

[30] Georgios Kavallieratos, Sokratis Katsikas, and Vasileios Gkioulos. 2019. Cyber-Attacks Against the Autonomous Ship. In *Computer Security*. Springer International Publishing, Cham, 20–36.

[31] Siddhartha Kumar Khaitan and James D McCalley. 2015. Design techniques and applications of cyberphysical systems: A survey. *IEEE Systems Journal* 9, 2 (2015), 350–365.

[32] Emrah Korkmaz, Andrey Dolgikh, Matthew Davis, and Victor Skormin. 2016. ICS security testbed with delay attack case study. In *Military Communications Conference, MILCOM 2016-2016 IEEE*. IEEE, 283–288.

[33] Emrah Korkmaz, Andrey Dolgikh, Matthew Davis, and Victor Skormin. 2016. Industrial Control Systems Security Testbed.

[34] Georgia Koutsandria, Reinhard Gentz, Mahdi Jamei, Anna Scaglione, Sean Peisert, and Chuck McParland. 2015. A real-time testbed environment for cyber-physical security on the power grid. In *Proceedings of the First ACM Workshop on Cyber-Physical Systems-Security and/or PrivaCy*. ACM, 67–78.

[35] H. Lin, A. Slagell, Z. T. Kalbarczyk, P. W. Sauer, and R. K. Iyer. 2018. Runtime Semantic Security Analysis to Detect and Mitigate Control-Related Attacks in Power Grids. *IEEE Transactions on Smart Grid* 9, 1 (Jan 2018), 163–178. https://doi.org/10.1109/TSG.2016.2547742

[36] Ren Liu, Ceeman Vellaithurai, Saugata S Biswas, Thoshitha T Gamage, and Anurag K Srivastava. 2015. Analyzing the cyber-physical impact of cyber events on the power grid. *IEEE Transactions on Smart Grid* 6, 5 (2015), 2444–2453.

[37] Guylaine M. Pollock, William Dee Atkins, Moses Schwartz, Adrian R. Chavez, Jorge Mario Urrea, Nicholas Pattengale, Michael James McDonald, Regis H. Cassidy, Ronald D. Halbgewachs, Bryan T. Richardson, and John C. Mulder. 2010. Modeling and simulation for cyber-physical system security research, development and applications. (01 2010). https://doi.org/10.2172/1028942

[38] Aditya P Mathur and Nils Ole Tippenhauer. 2016. SWaT: A water treatment testbed for research and training on ICS security. In *Cyber-physical Systems for Smart Water Networks (CySWater), 2016 International Workshop on*. IEEE, 31–36.

[39] Estefanía Etchevés Miciolino, Giuseppe Bernieri, Federica Pascucci, and Roberto Setola. 2015. Communications network analysis in a SCADA system testbed under cyber-attacks. In *Telecommunications Forum Telfor (TELFOR), 2015 23rd*. IEEE, 341–344.

[40] Thomas Morris, Anurag Srivastava, Bradley Reaves, Wei Gao, Kalyan Pavurapu, and Ram Reddi. 2011. A control system testbed to validate critical infrastructure protection concepts. *International Journal of Critical Infrastructure Protection* 4 (08 2011), 88–103. https://doi.org/10.1016/j.ijcip.2011.06.005

[41] Thomas Morris, Anurag Srivastava, Bradley Reaves, Wei Gao, Kalyan Pavurapu, and Ram Reddi. 2011. A control system testbed to validate critical infrastructure protection concepts. *International Journal of Critical Infrastructure Protection* 4, 2 (2011), 88–103.

[42] Igor Nai Fovino, Andrea Carcano, Marcelo Masera, and Alberto Trombetta. 2009. An experimental investigation of malware attacks on SCADA systems. *International Journal of Critical Infrastructure Protection* 2 (12 2009), 139–145. https://doi.org/10.1016/j.ijcip.2009.10.001

[43] Cuong Nguyen. 2018. NIST Smart Grid and CPS Newsletter-December 2017. (2018).

[44] NIST. 2010. Cyber Ranges. https://www.nist.gov/sites/default/files/documents/2018/02/13/cyber_ranges.pdf

[45] NIST. 2012. Guide for Conducting Risk Assessments - NFORMATION SECURITY.

[46] US Department of Defense. 2010. Reference Architecture Description. https://dodcio.defense.gov/Portals/0/Documents/DIEA/Ref_Archi_Description_Final_v1_18Jun10.pdf

[47] Shiva Poudel, Zhen Ni, and Naresh Malla. 2017. Real-time cyber physical system testbed for power system security and control. *International Journal of Electrical Power & Energy Systems* 90 (2017), 124–133.

[48] Ishaani Priyadarshini. 2018. *Features and Architecture of The Modern Cyber Range: A Qualitative Analysis and Survey*. Ph.D. Dissertation.

[49] Qais Qassim, Norziana Jamil, Izham Zainal Abidin, Mohd Ezanee Rusli, Salman Yussof, Roslan Ismail, Fairuz Abdullah, Norhamadi Ja'afar, Hafizah Che Hasan, and Maslina Daud. 2017. A Survey of SCADA Testbed Implementation Approaches. *Indian Journal of Science and Technology* 10, 26 (2017).

[50] Carlos Queiroz, Abdun Mahmood, and Zahir Tari. 2011. SCADASim a framework for building SCADA simulations. *IEEE Trans. Smart Grid* 2 (12 2011), 589–597. https://doi.org/10.1109/TSG.2011.2162432

[51] Bradley Reaves and Thomas Morris. 2012. An open virtual testbed for industrial control system security research. *International Journal of Information Security* 11, 4 (2012), 215–229.

[52] Luis Sanchez, Luis Munoz, Jose Antonio Galache, Pablo Sotres, Juan R. Santana, Veronica Gutierrez, Rajiv Ramdhany, Alex Gluhak, Srdjan Krco, Evangelos Theodoridis, and Dennis Pfisterer. 2014. SmartSantander: IoT experimentation over a smart city testbed. *Computer Networks* 61 (2014), 217 – 238. https://doi.org/10.1016/j.bjp.2013.12.020 Special issue on Future Internet Testbeds âĂŞ Part I.

[53] Christos Siaterlis and Genge Bela. 2014. Cyber-Physical Testbeds. *Commun. ACM* 57 (06 2014), 64–73. https://doi.org/10.1145/2602575

[54] Shachar Siboni, Vinay Sachidananda, Asaf Shabtai, and Yuval Elovici. 2016. Security Testbed for the Internet of Things. (10 2016).

[55] Ahnaf Siddiqi, Nils Ole Tippenhauer, Daisuke Mashima, and Binbin Chen. 2018. On practical threat scenario testing in an electric power ICS testbed. In *Proceedings of the 4th ACM Workshop on Cyber-Physical System Security*. ACM, 15–21.

[56] Prateek Singh, Saurabh Garg, Vinod Kumar, and Zia Saquib. 2015. A testbed for SCADA cyber security and intrusion detection. In *Cyber Security of Smart Cities, Industrial Control System and Communications (SSIC), 2015 International Conference on*. IEEE, 1–6.

[57] Kimberly Tam and Kevin Jones. 2018. Cyber-Risk Assessment for Autonomous Ships. https://doi.org/10.1109/CyberSecPODS.2018.8560690

[58] Eniye Tebekaemi and Duminda Wijesekera. 2016. Designing an IEC 61850 based power distribution substation simulation/emulation testbed for cyber-physical security studies. In *Proceedings of the First International Conference on Cyber-Technologies and Cyber-Systems*. 41–49.

[59] A. Tekeoglu and A. S. Tosun. 2016. A Testbed for Security and Privacy Analysis of IoT Devices. In *2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. 343–348. https://doi.org/10.1109/MASS.2016.051

[60] Abebe Tesfahun and Lalitha Bhaskari. 2016. A SCADA testbed for investigating cyber security vulnerabilities in critical infrastructures. *Automatic Control and Computer Sciences* 50 (01 2016), 54–62. https://doi.org/10.3103/S0146411616010090

[61] Jan Vykopal, Radek Oslejšek, Pavel Čeleda, Martin Vizvary, and Daniel Tovarňák. 2017. Kypo cyber range: Design and use cases. (2017).

[62] Evangelia Xypolytou, Joachim Fabini, Wolfgang Gawlik, and Tanja Zseby. 2017. The FUSE testbed: establishing a microgrid for smart grid security experiments. *e & i Elektrotechnik und Informationstechnik* 134, 1 (2017), 30–35.

[63] X. Zheng, L. Pan, H. Chen, R. D. Pietro, and L. Batten. 2017. A Testbed for Security Analysis of Modern Vehicle Systems. In *2017 IEEE Trustcom/BigDataSE/ICESS*. 1090–1095. https://doi.org/10.1109/Trustcom/BigDataSE/ICESS.2017.357

12 Article VIII: Cyber Risk Propagation and Optimal Selection of Cybersecurity Controls for Complex Cyberphysical Systems [8]

sensors

150

MDPI

Article

Cyber Risk Propagation and Optimal Selection of Cybersecurity Controls for Complex Cyberphysical Systems

Georgios Kavallieratos, Georgios Spathoulas and Sokratis Katsikas *

Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik N-2815, Norway; georgios.kavallieratos@ntnu.no (G.K.); georgios.spathoulas@ntnu.no (G.S.)
* Correspondence: sokratis.katsikas@ntnu.no; Tel.: +47-91138581

Abstract: The increasingly witnessed integration of information technology with operational technology leads to the formation of Cyber-Physical Systems (CPSs) that intertwine physical and cyber components and connect to each other to form systems-of-systems. This interconnection enables the offering of functionality beyond the combined offering of each individual component, but at the same time increases the cyber risk of the overall system, as such risk propagates between and aggregates at component systems. The complexity of the resulting systems-of-systems in many cases leads to difficulty in analyzing cyber risk. Additionally, the selection of cybersecurity controls that will effectively and efficiently treat the cyber risk is commonly performed manually, or at best with limited automated decision support. In this work, we propose a method for analyzing risk propagation and aggregation in complex CPSs utilizing the results of risk assessments of their individual constituents. Additionally, we propose a method employing evolutionary programming for automating the selection of an optimal set of cybersecurity controls out of a list of available controls, that will minimize the residual risk and the cost associated with the implementation of these measures. We illustrate the workings of the proposed methods by applying them to the navigational systems of two variants of the Cyber-Enabled Ship (C-ES), namely the autonomous ship and the remotely controlled ship. The results are sets of cybersecurity controls applied to those components of the overall system that have been identified in previous studies as the most vulnerable ones; such controls minimize the residual risk, while also minimizing the cost of implementation.

Keywords: cybersecurity; cyber physical systems; cyber risk propagation; cybersecurity controls; autonomous vessels

Citation: Kavallieratos, G.; Spathoulas, G.; Katsikas, S. Cyber Risk Propagation and Optimal Selection of Cybersecurity Controls for Complex Cyberphysical Systems. *Sensors* **2021**, *1*, 1. https://doi.org/

Academic Editor: Sherali Zeadally

Received: 24 January 2021
Accepted: 23 February 2021
Published:

1. Introduction

Cyber-Physical Systems (CPSs) are characterized by the strong coupling of the physical and the cyber worlds. The inevitable dependence on highly automated procedures and the increasing integration of *physical parts* to highly interconnected *cyber parts* render CPSs vulnerable to cyber attacks. On the other hand, the wide use of such systems in various critical domains [1] (e.g., Smart Grid, Intelligent Transportation Systems, Medical devices, Industrial Control Systems, etc.) increases the impact of such cyber attacks. Furthermore, the *System of Systems* (SoS) nature of interconnected, complex CPSs [2] introduces challenges in addressing security risks. In this context, a complex CPS comprises other CPSs that are interconnected, and control and information flows exist among them. These flows constitute pathways that a cyber attack may leverage to propagate from component to component. More specifically, both or either of the likelihood of the attack and its impact, if successful, may propagate. Because likelihood and impact are the constituents of risk, the cyber risk of the overall system is related to the individual cyber risk of each interconnected component. This in principle means that knowledge of the cyber risk of the individual components of a complex CPS may be leveraged to assess the cyber risk of the overall system, thus also facilitating the analysis of large scale, complex CPSs through a divide-and-conquer-like approach to cyber risk assessment.

The assessment of risk is one of the steps in the risk management process [3] that concludes with treating the risk by means of controls that aim at achieving retention, reduction, transfer, or avoidance of the risk [4]. In the general case, each risk can be treated by a number of possible cybersecurity controls, each of which with varying effectiveness and efficiency characteristics. Note that the same control may be effective and efficient in treating more than one risk. Therefore, an important task in formulating the risk treatment plan is the selection of the optimal set of cybersecurity controls, the criterion of optimality in this context being effectiveness and efficiency. Because of the complexity of formulating this as a formal optimization problem, particularly when there are more than one criteria of optimality, the selection of the cybersecurity controls is largely performed empirically, at best with some automated decision support.

In this paper, we propose a novel method for identifying a set of effective and efficient cybersecurity controls for large scale, complex CPSs comprising other CPSs as components. We also propose a method for assessing the aggregated risk that results by taking into account the risk of the individual components and the information and control flows among these components. Specifically, we leverage evolutionary computing to develop a cybersecurity control selection algorithm that uses the aggregated cyber risk of a complex CPS to generate a set of effective and efficient cybersecurity controls to reduce this risk. The algorithm selects the cybersecurity controls among the list of such controls in the NIST Guidelines for Industrial Control Systems Security [5]. We illustrate the workings of the proposed method by applying it to the navigational systems of two instances of the Cyber-Enabled Ship (C-ES), i.e., vessels with enhanced monitoring, communication, and connection capabilities that include remotely controlled and fully autonomous ships [6]. The C-ES comprises a variety of interconnected and interdependent CPSs [7], and, as such, it constitutes a complex CPS. Specifically, we derive the set of cybersecurity controls for both the autonomous and the remotely controlled vessel.

Thus, the contribution of this work is as follows:

- A novel method for assessing the aggregate cybersecurity risk of a large scale, complex CPS comprising components connected via links that implement both information and control flows, by using risk measures of its individual components and the information and control flows among these components.
- A novel method for selecting a set of effective and efficient cybersecurity controls among those in an established knowledge base, that reduce the residual risk, while at the same time minimizing the cost.
- Sets of cybersecurity controls for the navigational systems of two instances of the C-ES, namely the remotely controlled ship and the autonomous ship, derived by employing the two methods.

The remainder of this paper is structured as follows: Section 2 reviews the related work in the areas of cyber risk propagation and aggregation; optimal selection of cybersecurity controls; and C-ES risk management. Section 3 provides the background knowledge on genetic algorithms, and on the STRIDE (*S*poofing, *T*ampering, *R*epudiation, *I*nformation disclosure, *D*enial of Service, and *E*levation) and DREAD (*D*amage, *R*eproducibility, *E*xploitability, *A*ffected, and *D*iscoverability) risk assessment methods that is necessary to make the paper self-sustained. Sections 4 and 5 present the proposed method for risk aggregation in complex CPSs and the proposed method for optimal cybersecurity control selection, respectively. In Section 6, we apply the proposed methods to the remotely controlled and the autonomous ship cases and discuss the results. Finally, Section 7 summarizes our conclusions and outlines topics for future research work.

2. Related Work

Cyber risk is evaluated as a function of the likelihood of an adverse event, such as an attack, occurring; and of the impact that will result when the event occurs. In order for an adverse event to occur, a threat has to successfully exploit one or more vulnerabilities; this can be done by launching one of a number of possible attacks. Hence, the likelihood

of the event occurring is, in turn, determined by the likelihood of the threat successfully exploiting at least one vulnerability. Accordingly, in order to analyze how the cyber risk propagates in a complex system made up by interconnected components that are systems by themselves requires analyzing how both the likelihood of the event and its impact propagates. Once this analysis is accomplished, the aggregate cyber risk of the complex system can be assessed.

Several security risk assessment methods applicable to general purpose IT systems have appeared in the literature (see Reference [8] for a comprehensive survey). Even though several of these methods can be and have been applied to CPSs, they cannot accurately assess cyber risks related to CPSs according to Reference [9], where a number of approaches for risk assessment for CPSs are listed. A review of risk assessment methods for CPSs, from the perspective of safety, security, and their integration, including a proposal for some classification criteria was made in Reference [10]. A survey of IoT-enabled cyberattacks that includes a part focused on CPS-based environments can be found in Reference [11]. Cyber risk assessment methods for CPSs more often than not are domain specific, as they need to take into account safety as an impact factor additional to the "traditional" impact factors of confidentiality, integrity, and availability. For example, an overview of such methods specific to the smart grid case is provided in Reference [12]. A review of the traditional cybersecurity risk assessment methods that have been used in the maritime domain, is provided in Reference [13]. Additionally, various risk assessment methods have been proposed to analyze cyber risk in autonomous vessels [14–16].

Several works in the literature have studied how individual elements of cyber risk propagate in a network of interconnected systems; both deterministic and stochastic approaches have been used to this end. A threat likelihood propagation model for information systems based on the Markov process was proposed in Reference [17]. An approach for determining the propagation of the design faults of an information system by means of a probabilistic method was proposed in Reference [18]. A security risk analysis model (SRAM) that allows the analysis of the propagation of vulnerabilities in information systems, based on a Bayesian network, was proposed in Reference [19]. Methods for evaluating the propagation of the impact of cyber attacks in CPSs have been proposed in Reference [20–22], among others. Epidemic models were initially used to study malware propagation in information systems [17]. The propagation of cybersecurity incidents in a CPS is viewed as an epidemic outbreak in Reference [23] and is analyzed using percolation theory. The method was shown to be applicable for studying malware infection incidents, but it is questionable whether the epidemic outbreak model fits other types of incidents. Percolation theory was also used in Reference [24] to analyze the propagation of node failures in a network of CPSs comprising cyber and physical nodes organized in two distinct layers, such as in the case of the power grid. The Susceptible–Exposed–Infected–Recovered (SEIR) infectious disease model was used in Reference [25] to study malware infection propagation in the smart grid. A quantitative risk assessment model that provides asset-wise and overall risks for a given CPS and also considers risk propagation among dependent nodes was proposed in Reference [26].

A method for assessing the aggregate risk of a set of interdependent critical infrastructures was proposed in Reference [27,28]. The method provides an aggregate cyber risk value at the infrastructure level, rather than a detailed cyber risk assessment at the system/component level. Thus, it is suitable for evaluating the criticality of infrastructure sectors, but not for designing cybersecurity architectures or for selecting appropriate cybersecurity controls. A similar approach for the Energy Internet [29] was followed to develop an information security risk algorithm based on dynamic risk propagation in Reference [30]. A framework for modeling and evaluating the aggregate risk of user activity patterns in social networks was proposed in Reference [31]. A two-level hierarchical model was used in Reference [32] to represent the structure of essential services in the national cyberspace, and to evaluate the national level (aggregate) risk assessment by taking into account cyber threats and vulnerabilities identified at the lower level.

Based on the above discussion, it is evident that the problem of risk propagation and risk aggregation for complex systems, on one hand, and the problem of optimal selection of cybersecurity controls, on the other, have been individually studied. The conjunct problem of identifying the optimal set of cybersecurity controls that reduces the aggregate risk in a complex CPS cannot be approached by sequential application of methods each of which addresses the problem's components, due to the inherent nonlinearity of the risk propagation, risk aggregation, and control selection processes on one hand, and the intertwining of these processes. To the best of our knowledge, no method that solves this conjunct problem is currently available.

On the other hand, the systematic selection of cybersecurity controls has been mostly examined in the literature in attempting to identify the optimal set of controls for IT systems within a specified budget; examples of such approaches are those in Reference [33–35]. The outline of a programming tool that supports the selection of countermeasures to secure an infrastructure represented as a hierarchy of components was provided in Reference [36]. A methodology based on an attack surface model to identify the countermeasures against multiple cyberattacks that optimize the Return On Response Investment (RORI) measure is proposed in Reference [37]. However, to the best of our knowledge, a method that selects a set of cybersecurity controls that simultaneously optimizes both effectiveness and efficiency, by minimizing the residual risk and the cost of implementation, is still to be proposed.

The work described in this paper addresses these research gaps.

3. Background

3.1. Evolutionary/Genetic Algorithms

Genetic algorithms (GAs) are randomized search algorithms that imitate the structures of natural genetics and the mechanisms of natural selection [38]. They imitate biological genomes by means of strings structures that represent individuals and are composed of characters belonging to a sepcified alphabet. These structures form populations that evolve in time by means of a randomized exchange scheme that implements the principle of survival of the fittest; in every new generation, a new set of individuals is created, using parts of the fittest members of the old set, whilst also possibly retaining some of the fittest members of the old generation. GAs can be very useful when it comes to problems with very large solution spaces, where it is infeasible to exhaustively search the solution space. It should, however, be noted that GAs are not guaranteed to find the global optimum solution to a problem; however, they do find "acceptably good" solutions.

For designing a GA, a *coding scheme* that codes the parameter space; a set of *operators* to be used to each generation to generate the next generation; and a *fitness function* that measures the fitness of each individual as a functional of the function that we are trying to optimize need to be defined. The coding scheme and the fitness function to be used depend on the characteristics of the optimization problem on which the GA will be applied. However, a commonly used coding scheme is to use the binary alphabet to represent each element (gene) in a string (genome). On the other hand, the most commonly used operators are the *reproduction* operator, the *crossover* operator, and the *mutation* operator. These have been found to be both computationally simple and effective in a number of optimization problems [39].

The operators are used to evolve populations by creating new individuals that will form the new generation. To this end, the reproduction operator tentatively selects individuals with high fitness function values as candidate parents for the next generation, by means of a randomized technique, such as a *roulette wheel selection scheme*. The selected parents may mate by means of the crossover operator, that randomly selects pairs of mates and creates new individuals, by combining elements of both parents, these elements being selected at random. As in biological populations, random genetic alterations (mutations) sometimes result in genetically fitter individuals. Such alterations, that happen with small probability, are implemented in GAs by means of the mutation operator.

The generic GA addresses unconstrained optimization problems. However, constrained optimization problems are encountered more often than not, including the problem addressed in this work, as will be seen in the sequel. Constraints can be modeled as either equality relations, that can be incorporated within the function to be optimized; or as inequality relations, that may be handled either by simply evaluating the fitness of each individual and then check to see whether any constraints are violated, or by employing a penalty method. In the former (reactive) strategy, if an individual violates a constraint, it is assigned a fitness value equal to zero. In the latter (proactive) strategy, the fitness of an individual that violates a constraint is decreased by an amount proportional to the cost of the violation.

3.2. STRIDE

STRIDE [40] is a cyber security threat modeling method that was developed at Microsoft in 1999. It facilitates the process of identifying and analyzing six types of threats, namely *S*poofing, *T*ampering, *R*epudiation, *I*nformation disclosure, *D*enial of Service, and *E*levation of privileges, in which the initials form the acronym *STRIDE*. Each of these threats corresponds to the violation of a desirable property (security objective) of the system under study, as follows:

- Spoofing corresponds to violation of authenticity;
- Tampering corresponds to violation of integrity;
- Repudiation corresponds to violation of non-repudiability;
- Information disclosure corresponds to violation of confidentiality;
- Denial of service corresponds to violation of availability; and
- Elevation of privileges corresponds to violation of authorization.

STRIDE can be used to analyze threats for systems being in a variety of development phases, even for systems at the design phase; thus, it enables adherence to security-by-design principles [41]. Furthermore, even though originally designed for software systems, STRIDE has been also used in ecosystem environments where CPSs are prominently present [42–44]. In particular, a modified version of STRIDE was proposed and used in Reference [6] to model threats, to develop cyber attack scenarios, and to qualitatively assess the accordant risks for a number of CPSs in the C-ES ecosystem.

3.3. DREAD

DREAD is a security risk assessment model that, like STRIDE, was developed as part of Microsoft's threat modeling and risk analysis process. The name is an acronym made up from the initials of the characteristics of the risk associated with each attack scenario being analyzed, namely *D*amage (what is the extent of the damage that the attack is expected to inflict on the system); *R*eproducibility (how easy it is to reproduce the attack); *E*xploitability (the extent of the resources that the adversary needs to launch the attack); *A*ffected users/systems (how many people and/or systems will be affected); and *D*iscoverability (how easy is it for the adversary to identify vulnerabilities to exploit for launching the attack) [45].

STRIDE and DREAD are interrelated: the former allows the qualitative security analysis of the system, whilst the latter quantifies the identified risks. According to the approach in Reference [22], the values (High, Medium, Low) of the DREAD variables associated with each STRIDE threat $t \in \{S, T, R, I, D, E\}$ are determined by applying a specific set of criteria, shown in Table 1; these have been adapted from those in Reference [45], so as to include CPS aspects, and are further analyzed in Reference [22].

Table 1. Criteria for determining the values of the DREAD (*D*amage, *R*eproducibility, *E*xploitability, *A*ffected, and *D*iscoverability) variables [22,44].

	High (3)	Medium (2)	Low (1)
D	The adversary is able to bypass security mechanisms; get administrator access; upload/modify the CPS content.	Leakage of confidential information of the CPSs (functions/source code); partial malfunction/disruption of the system.	Leakage of non-sensitive information; the attack is not possible to extend to other CPSs on-board.
R	The attack can be reproduced at anytime.	The adversary is able to reproduce the attack, but under specific risk conditions.	Although the attacker knows the CPS's vulnerabilities/faults, they are unable to launch the attack.
E	The attack can be performed by a novice adversary, in a short time.	A skilled adversary may launch the attack.	The attack requires an extremely skilled person and in-depth knowledge of the targeted CPS.
A	All CPSs are affected.	Some users/systems, with non-default configuration are affected.	The attack affects only the targeted CPS.
D	The CPS's vulnerabilities are well known, and the attacker is able to access the relevant information to exploit them.	The CPS's vulnerabilities/faults are not well known and the adversary needs to access the CPS.	The threat has been identified, and the vulnerabilities have been patched.

Then, the risk value R_t^s associated with each STRIDE threat $t \in \{S, T, R, I, D, E\}$ for system s is calculated by using the following formulas [41,44,45]:

$$Impact_t^s = \frac{Damage + Affectedsystems}{2}, \tag{1}$$

$$Likelihood_t^s = \frac{Reproducibility + Exploitability + Discoverability)}{3}, \tag{2}$$

$$Risk_t^s = \frac{(Impact_t^s + Likelihood_t^s)}{2}. \tag{3}$$

$Impact_t^s$ represents a measure of the effect a successful attack materializing threat t has on the component s; $Likelihood_t^s$ represents a measure of how likely it is for threat t to materialize on s.

Both STRIDE and DREAD have been used in Reference [44] to assess the cyber risk of Cyber-Physical Systems (CPSs) on board the C-ES paradigm.

4. Cyber Risk Propagation and Aggregation

4.1. System Model

Assume a CPS consisting of N interconnected components, each denoted by c_i, $i = 1, ...N$. This system can be represented by a directed graph of $N + 1$ nodes, the system itself being one of the nodes, denoted as c_0. The edges of the graph represent information and control flows between the nodes. An edge from node A to node B indicates the existence of either an information flow or a control flow, from A to B. A consequence of the existence of such an edge is that a cybersecurity event at node A affects node B, as well. For example, in the simple graph of Figure 1, a cybersecurity event at node A will have effect on node B, as well, while a cybersecurity event at node B will have effect on both nodes A and C. The relationship "has effect" can be quantified by assigning an *effect coefficient* to each flow.

These are denoted henceforth by eff^a_{AB}, where $a = I$ for the information flow, and $a = C$ for the control flow, respectively. One way of assigning values to these coefficients is to use the inverse of the *in degree centrality*, i.e., the number of flows arriving to that node, denoted by IDC. Following this approach, the case in which information arrives to node B only through node A, will result in a much higher eff^I_{AB} than the case where information arrives to node B from a large number of nodes, including A. By definition, the values of all effect coefficients lie in the [0, 1] range and provide an indication of the percentage of the damage that is propagated from one node to the other. The *total effect coefficient* eff^T_{AB} is computed as a function of eff^I_{AB} and eff^C_{AB}, as in Equation (4).

The function f in Equation (4) has to be instantiated according to the requirements of the domain to which the methodology is applied and/or to specific characteristics of components A and B with regards to the criticality of information and control flows between them. For example, one option is to select f as the average of the effect coefficients. This option reflects equal importance of the information and the control flows in risk propagation, and it has been used in the illustrative application of the method presented in Section 6.

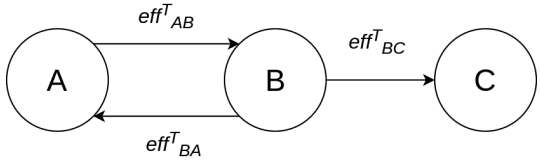


Figure 1. Effect relationship between nodes.

$$eff^T_{AB} = f(eff^I_{AB}, eff^C_{AB}), \tag{4}$$

where $eff^I_{AB} = \frac{1}{IDC^I_B}$, $eff^C_{AB} = \frac{1}{IDC^C_B}$.

Another example is that of a cyber-physical system that mainly aims at sensing and processing data coming from a process, e.g., an electric power smart meter. In such systems, information workflows are more significant than control flows, and a function f of the form $eff^T_{AB} = a * eff^I_{AB} + b * eff^C_{AB}$ with $a + b = 1, a > b$ would be a good choice. On the other hand, for a cyber-physical system that aims at controlling a process, e.g., a smart grid digital switch, a variant of the same function f but with $a + b = 1, b > a$ would be more appropriate, as control flows are more likely to enable cyber risk propagation between components.

4.2. Aggregate Risk

For any threat t, the *aggregate* risk $R_t^{agg_{c_j}}$ of component c_j is (applying the worst case scenario principle [28]) given by:

$$R_t^{agg_{c_j}} = \max(R_t^{dir_{c_j}}, R_t^{prop_{c_j}}), \tag{5}$$

where $R_t^{dir_{c_j}}$ (*direct* risk) is the risk when c_j is not connected to any other component c_k, $k \neq j$, which is calculated by means of Equations (1)–(3), and $R_t^{prop_{c_j}}$ (*propagated* risk) is the risk that c_j faces because of its connections to other components. These connections may be over any, possibly multi-hop, path p_l from any node k to j, $k \neq j$. Applying again the worst case scenario principle, $R_t^{prop_{c_j}}$ is calculated as:

$$R_t^{prop_{c_j}} = \max_{p_l} R_t^{prop_{c_j}^{p_l}}, \tag{6}$$

where $R_t^{prop_{c_j}^{p_l}}$ is the risk of component c_j associated with threat t and propagated along path p_l.

When a threat materializes against component c_i, it will also create an effect to component c_j, if c_i and c_j are connected. In the absence of controls, the likelihood that this will happen is equal to the likelihood that the threat will materialize against c_i in the first place. In contrast, the impact that this event has on c_j is only a fraction of the impact the event has on any c_k on any path p_l from c_i to c_j. This fraction is represented by $eff_{p_l}^T$ and is calculated by

$$eff_{p_l}^T = \prod_{i=1}^{j-1} eff_{c_i c_{i+1}}^T. \tag{7}$$

Accordingly, the risk propagated over path p_l, originating at component (node) c_i and terminating at component (node) c_j, is calculated by:

$$R_t^{prop_{c_j}^{p_l}} = \frac{eff_{c_i c_j}^{T_{p_l}} * Impact_t^{c_i} + L_t^{c_i}}{2}. \tag{8}$$

The system as a whole is represented by c_0; therefore, the (global) risk of threat t for the system is given by:

$$R_t^s = R_t^{agg_{c_0}} = \max(R_t^{dir_{c_0}}, R_t^{prop_{c_0}}), \tag{9}$$

where the direct risk for the system is not applicable ($R_t^{dir_{c_0}} = 0$) and the propagated risk for the system is calculated as for any other node ($R_t^{prop_{c_0}} = \max_{p_l} R_t^{prop_{c_0}^{p_l}}$), thus

$$R_t^s = \max_{p_l} R_t^{prop_{c_0}^{p_l}} \tag{10}$$

In order to showcase how the global risk calculation works and also to shed light on an underlying subtle assumption, consider the example system shown in Figure 2. In order to calculate the aggregate risk of each c_i, $i = 1, 2, 3$, we need to calculate the propagated risks, and this requires identifying all possible paths originating at any node and terminating at c_i, $i = 1, 2, 3$, respectively. The propagated risk for c_3 is equal to zero, as there is no such path. Nodes c_1 and c_2 are interconnected; therefore, a loop exists between them. Consequently, if we allow circular paths to be considered, there are infinite paths between these two nodes, and the computation in Equation (7) would be endless. However, by noticing that the value of the total effect coefficient becomes, by definition, negligible after a couple of hops, we are able to disregard circular paths in its computation.

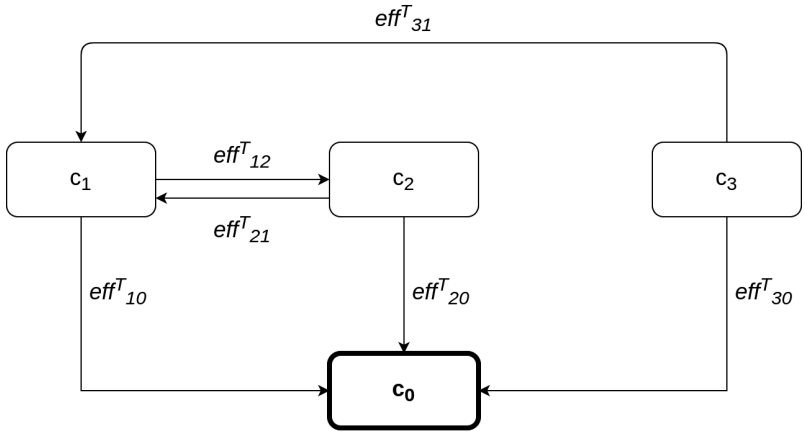


Figure 2. An example of a system.

Therefore, the global risk of a system can be calculated by the algorithm in Algorithm 1. As can be seen in Algorithm 1, nodes along a path are processed recursively, starting at the end of the path. If a node is already in the path, it is not included again, so as to avoid cyclic paths.

Algorithm 1: Global system risk calculation algorithm.

Result: Global system risk is calculated as R_t^s
Function `process_node`(c_j, eff, p_l)**:**
  $L = L_t^{c_j}$;
  $I = I_t^{c_j}$;
  $R = \frac{L+I}{2}$;
  foreach *edge from* c_i *to* c_j **do**
    if $c_i \notin p_l$ **then**
      $p_l = p_l \cup \{c_i\}$;
      $L', I' = process\_node(c_i, eff_{c_i c_j}, p_l)$;
      $R' = \frac{L'+I'}{2}$;
      if $R' > R$ **then**
        $L = L'$;
        $I = I'$;
        $R = R'$;
      end
    end
  end
  return $eff * L, I$;
L, I=`process_node`(c_0, *1*, $\{c_0\}$);
$R_t^s = \frac{L+I}{2}$;

5. Optimal Cybersecurity Control Selection

5.1. Cybersecurity Controls

We assume that there exists a list of controls available to apply to the components of the system. Each control m, when applied to component c_i, has a potential effect on the values of $Impact_t^{c_i}$ and $Likelihood_t^{c_i}$ that are used in the calculation of the cyber risk, such effect depending on the effectiveness and the nature of the control. We denote the new

Likelihood and Impact values of threat t that result after the application of control m to c_i by $Likelihood_{t,m}^{c_i}$ and $Impact_{t,m}^{c_i}$, respectively. These values can be calculated by re-applying DREAD to the system, which is now protected by m.

Additionally, for each control m, a cost metric $Cost_m$ is defined. This metric is expressed on a 1–5 scale, corresponding to the qualitative classifications very low cost, low cost, medium cost, high cost, and very high cost. Note that the use of this scale was dictated by the fact that it is difficult to measure the cost of implementing a control. However, if such a measure is available, the replacement of the value in the 1–5 scale with the actual cost of the control is straightforward.

For a system with N components and a list with M controls with the cost metrics vector $C = [cost_1, cost_2, ..., cost_M]$, the following binary matrix AC compactly depicts the applied controls throughout the system:

$$AC = \begin{bmatrix} ac_{1,1} & ac_{1,2} & ... & ac_{1,N} \\ ac_{2,1} & ac_{2,2} & ... & ac_{2,N} \\ ... & ... & ... & ... \\ ac_{M,1} & ac_{M,2} & ... & ac_{M,N} \end{bmatrix}, \tag{11}$$

where

$$ac_{i,j} = \begin{cases} 0, \textit{if control i is not applied to component j} \\ 1, \textit{if control i is applied to component j} \end{cases}. \tag{12}$$

Then, the total cost TC_{AC} of the applied controls solution AC is given by $TC_{AC} = AC * C$.

5.2. Optimization Method

The optimization problem to be solved is to select the optimal (effective and efficient) set of controls among a list of possible ones. This amounts to selecting the set of controls AC that minimizes the system *residual* risk $R_{t_{AC}}^s$, at the lowest total cost TC. A closed formula that would allow the application of an exact optimization method, and thus the calculation of the globally optimum solution to the problem, is not possible to construct, unless many, not necessarily realistic, assumptions are made. On the other hand, the large size of the search space (all candidate solutions) prohibits the exhaustive search approach. Hence, a heuristic optimization method has to be employed [46]; we have selected to use a genetic algorithm, even though any other heuristic optimization method would, in principle, be applicable.

The design parameters of the genetic algorithm are as follows:

- The search space comprises all possible combinations of controls applied to components.
- Each individual solution is represented by the matrix AC, which is transformed into a binary vector of size $M * N$. The value of each element of the vector represents the decision to apply a specific control to a specific component or not. For example, for a system with three components and two controls, the solution would be denoted by the vector $[ac_{11}, ac_{21}, ac_{12}, ac_{22}, ac_{13}, ac_{23}]$, assuming that all controls are applicable to all components.
- The fitness function is defined as $fit(AC) = R_{t_{AC}}^s + C_{norm}(AC)$, where $C_{norm}(AC) = \frac{TC_{AC}}{TC_{max}}$, with TC_{max} being the largest possible cost, that results when applying all available controls to all system components.
- The initial population size is 100.
- The mutation probability is 0.1.
- The next generation is determined by uniform crossover, with crossover probability equal to 0.5, an elite ratio of 0.01, and 0.3 of the population consisting of the fittest members of the previous generation (aka parents).
- The algorithm terminates when the maximum number of allowed iterations is used. This number is calculated as $iter_{max} = 50 * \sum_{i=1,j=1}^{i=M,j=N} ac_{ij}$.

The algorithm for selecting the optimal set of security controls is depicted in Algorithm 2.

Note that the fitness function consists of two elements, namely the residual risk (which takes values in [0, 3]) and the normalized cost (which takes values in [0, 1]). This non-symmetric approach has been selected to put emphasis on the importance of reducing the residual risk, even by bearing larger cost. This approach results in initial iterations of the algorithm tending to generate solutions that minimize the residual risk. In later iterations of the algorithm, the less costly combinations of controls prevail, among those that lead to the maximum possible risk reduction.

6. Application to the C-ES

Autonomous and remotely controlled ships—both variants of the Cyber-Enabled Ship (C-ES)—are being increasingly developed. At the same time, the maritime transportation sector contributes significantly to the gross domestic product of many countries around the world. It is not surprising, then, that the cybersecurity of the sector has been designated a very high priority by international organizations [47] and national governments [48] alike. The CPSs comprising the C-ES were identified, and the overall ICT architecture of the C-ES in the form of a tree structure was proposed in Reference [6]. An extended Maritime Architectural Framework (e-MAF) was proposed, and the interconnections, dependencies, and interdependencies among the CPSs of the C-ES were described in Reference [7]. These results are depicted in the form of directed graphs in Figures 3–6 for the two variants of the C-ES. Furthermore, an initial threat analysis of the generic ICT architecture of the C-ES identified the three most vulnerable onboard systems, namely the Automatic Identification System (AIS), the Electronic Chart Display Information System (ECDIS), and the Global Maritime Distress and Safety System (GMDSS) [6]. These results were verified by means of the comprehensive threat and risk analysis that was presented in Reference [44]. The most critical attack paths within the navigational CPSs of the C-ES were identified in Reference [22]. The cybersecurity and safety requirements for the CPSs of the C-ES were identified in Reference [49,50], and an initial set of cybersecurity controls that satisfy these requirements was proposed in Reference [44].

Building upon earlier work, and as a step towards defining the cybersecurity architecture of such vessels, we selected the CPSs of the C-ES to illustrate the applicability of the methods proposed in this paper. The results are presented in the sequel for the autonomous and the remotely controlled vessel.

6.1. The Cyber-Enabled Ship

The CPSs of the C-ES were identified and described in Reference [6], where a threat analysis and a qualitative risk analysis were carried out, and the most vulnerable onboard systems were identified. Three distinct sub-groups of onboard CPSs were identified, namely the bridge CPSs; the engine CPSs; and the Shore Control Center (SCC) CPSs. The SCC is a sub-component of the remotely controlled vessel, that aims to control and navigate one or more ships from the shore. The interconnections, dependencies, and interdependencies of these CPSs were identified in Reference [7] and were later used to define the cybersecurity requirements of the C-ES in Reference [49]. The CPSs considered herein are:

- The Autonomous Navigation System (ANS) is responsible for the navigational functions of the vessel. ANS controls all the navigational sub-systems and communicates with the SCC by transmitting dynamic, voyage, static, and safety data to ensure the vessel's safe navigation.
- The Autonomous Ship Control (ASC) acts as an additional control for the C-ES and aims to assess the data derived from the sensors and from the SCC.
- The Advanced Sensor Module (ASM) automatically analyzes sensor data to enhance the environmental observations, such as ships in the vicinity. By leveraging sensor fusion techniques, this module analyzes data derived from navigational sensors, such as the Automatic Identification System (AIS) and the Radar.

Algorithm 2: Algorithm for selecting the optimal set of security controls

Result: Optimal set of security controls is identified

Function `calc_fitness`(*control_sets*)**:**
- control_sets_fit_scores = [];
- **foreach** *c in control_sets* **do**
  - control_sets_fit_scores[c] = fit_score(*c*);
- **end**
- **return** control_sets_fit_scores;

Function `select_parents`(*control_sets*,*control_sets_fit_scores*)**:**
- parents_control_sets = [] ;
- **foreach** *c in control_sets* **do**
  - **if** *control_sets_fit_scores[c] ∈ upper 30% of control_sets_fit_scores* **then**
    - parents_control_sets ← *c*;
  - **end**
- **end**
- **return** parents_control_sets;

Function `select_elite`(*control_sets*,*control_sets_fit_scores*)**:**
- elite_control_sets = [] ;
- **foreach** *c in control_sets* **do**
  - **if** *control_sets_fit_scores[c] ∈ upper 1% of control_sets_fit_scores* **then**
    - elite_control_sets ← *c*;
  - **end**
- **end**
- **return** elite_control_sets;

Function `crossover`(*parent_control_sets*)**:**
- control_sets = parent_control_sets;
- pop = |control_sets|;
- **while** *pop<100* **do**
  - $parent_a = random(parent\_control\_sets)$;
  - $parent_b = random(parent\_control\_sets)$;
  - $control\_set_{new} = crossover(parent_a, parent_b)$;
  - $control\_sets \leftarrow control\_set_{new}$;
  - $pop = pop + 1$;
- **end**
- **return** control_sets;

Function `mutation`(*control_sets*,*elite_control_sets*)**:**
- mutated_control_sets = [];
- **foreach** *c in control_sets* **do**
  - **if** *c ∈ elite_control_sets* **then**
    - *mutated_control_sets* ← *c*;
  - **else**
    - *mut_c* = *mutate*(*c*);
    - *mutated_control_sets* ← *mut_c*;
  - **end**
- **end**
- **return** *mutated_control_sets*;

Function `find_solution()`**:**
- $iter_{max} = 50 * \sum_{i=1,j=1}^{i=M,j=N} ac_{ij}$;
- $iter = 0$;
- control_sets ← 100 random sets;
- **while** $iter < iter_{max}$ **do**
  - control_sets_fit_scores = `calc_fitness`(*control_sets*);
  - parents_control_sets = `select_parents`(*control_sets*,*control_sets_fit_scores*);
  - elite_control_sets = `select_elite`(*control_sets*,*control_sets_fit_scores*);
  - control_sets = `crossover`(*parents_control_sets*);
  - control_sets = `mutation`(*control_sets*);
  - $iter = iter + 1$
- **end**
- fittest_control_set = fittest *c* ∈ *control_sets* **return** fittest_control_set;

`find_solution()`

Table 2. Impact values.

	Impact									
	ANS	**ASC**	**ASM**	**AIS**	**CA**	**ECDIS**	**SCC**	**RADAR**	**AP**	**VDR**
S	2.5	3	2.5	2	2.5	2.5	2.5	2.5	2	2
T	2.5	2	1.28	2.5	2.5	2	2	2.5	2.5	2
R	2	2.5	1.5	2	1.5	1.5	1.5	2	1.5	1.5
I	2.5	2.5	2	2	1.5	3	1.5	1	2	2
D	2.5	2.5	2	2	2.5	3	2.5	2	2.5	2
E	3	3	1.5	2.5	1.5	3	1.5	2	2	2

Table 3. Likelihood values.

	Likelihood									
	ANS	**ASC**	**ASM**	**AIS**	**CA**	**ECDIS**	**SCC**	**RADAR**	**AP**	**VDR**
S	1.33	1.33	2	2.66	1.33	2.32	1.66	2	1	1
T	1.33	2	1.28	2.33	1.66	2.33	1.33	1.66	1	1
R	1	1	1	2.66	1	1	1.33	1.33	1	1
I	1	1	1.33	2.66	1.33	1.66	1.33	1	1	1
D	1.33	1.66	2	2	1.33	2	1.66	2	1	1
E	1.33	1	1	1.33	1	1.66	1	1	1	1

- The Automatic Identification System (AIS) facilitates the identification, monitoring, and locating of the vessel by analyzing voyage, dynamic, and static data. Further, the AIS contributes to the vessel's collision avoidance system by providing real time data.
- The Collision Avoidance (CA) system ensures the safe passage of the vessel by avoiding potential obstacles. The system analyzes the voyage path by leveraging anti-collision algorithms conforming to the accordant COLREGs regulations [51].
- The Electronic Chart Display Information System (ECDIS) supports the vessel's navigation by providing the necessary nautical charts, along with vessel's attributes, such as position and speed.
- The marine RADAR provides the bearing and distance of objects in the vicinity of the vessel, for collision avoidance and navigation at sea.
- The Voyage Data Recorder (VDR) gathers and stores all the navigational data of the vessel specifically related to vessel's condition, position, movements, and communication recordings.
- The Auto Pilot (AP) controls the trajectory of the vessel without requiring continuous manual control by a human operator.

The methods proposed in Sections 4 and 5 used as input prior results, namely the system components and their interconnections that make up the system graph representation; the impact and likelihood values associated with the STRIDE threats and computed by means of DREAD for each individual component; and the list of available cybersecurity controls, along with information on their cost and effectiveness. Figures 3–6 depict the graph representations of the onboard navigational CPSs of the autonomous and of the remotely controlled ship, respectively, along with their interconnections and interdependencies [6,22,44]. Impact and likelihood values associated with the STRIDE threats and computed by means of DREAD are depicted in Tables 2 and 3 [44]. Each line of Tables 2 and 3 represents one of the STRIDE threats, indicated by the corresponding initial. Each column of the Table represents individual CPSs, indicated by their corresponding initials, as defined in Section 6.1. The values inside the cells are the corresponding impact (left table) and likelihood (right table) values per STRIDE threat and per individual component; these have been calculated by means of Equations (1) and (2), respectively. These values are subsequently used as input to Algorithm 1, to calculate the aggregate risk of each CPS.

The list of available cybersecurity controls has been defined based on the NIST guidelines for Industrial Control Systems security [5] by following a systematic process proposed in Reference [44]. The effectiveness and the cost of each security control are estimated considering their applicability, the extent to which each control reduces the impact or/and the likelihood, and the resources needed to implement it.

6.2. Optimal Controls for the Autonomous Ship

Autonomous ships are equipped with advanced interconnected CPSs able to navigate and sail the vessels without human intervention. The onboard navigational CPSs of the autonomous ship are described by the directed graphs $G_I(V, E)$ and $G_C(V, E)$ depicted in Figures 3 and 4, respectively, as discussed in detail in Reference [6,44]. $G_I(V, E)$ represents information flow connections and $G_C(V, E)$ control flow connections. Table 4 depicts the effect coefficients between all the considered systems. Each line and each column of Table 4 represents a CPS of the C-ES, indicated by their corresponding initials, as defined in Section 6.1 above. The values inside the cells are the effect coefficients between each pair of these systems; specifically, the value in the cell at row i and column j is the value of eff_{ij}^{T}. These have been calculated by means of Equation (13), which derives from Equation (4) when the function f is the average of the information and control effect coefficients. These values are also subsequently used as input to Algorithm 1, to calculate the aggregate risk of each CPS.

$$eff_{AB}^{T} = \frac{eff_{AB}^{I} + eff_{AB}^{C}}{2}. \tag{13}$$

It is worth noticing that CPSs with high information and control flows, such as the ANS and the ASC, are characterized by high values of the effect coefficient.

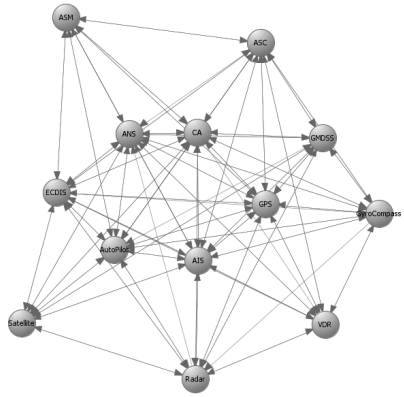


Figure 3. Autonomous ship—Navigational Cyber-Physical Systems (CPSs)–$G_I(V, E)$–Information flow connections.

Table 4. Effect coefficients—Autonomous ship.

	C-ES	AIS	ECDIS	VDR	ASM	RADAR	AP	CA	ANS	ASC
C-ES	0	0	0	0	0	0	0	0	0	0
ANS	0.208	0.208	0.208	0.208	0.208	0.166	0.208	0.208	0	0
ASC	0.055	0.055	0.055	0.055	0.055	0	0.055	0.055	0.055	0
ASM	0.321	0.071	0.071	0	0	0	0.071	0.321	0.071	0.071
AIS	0.041	0	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041
CA	0.211	0.211	0.045	0.045	0.045	0.045	0.045	0	0.211	0.045
ECDIS	0.05	0.05	0	0.05	0.05	0.05	0.05	0.05	0.05	0.05
RADAR	0,055	0	0.055	0.055	0	0	0.055	0.055	0.555	0
AP	0,045	0.045	0.045	0	0	0.045	0	0.045	0.045	0.045
VDR	0,062	0.062	0.062	0	0	0.062	0	0	0.062	0.062

The security controls in the optimal set are selected from the initial list of available controls by applying the method described in Section 5. Table 5 depicts the optimal set of security controls per STRIDE threat and per CPS component. It also depicts the associated initial global risk (without controls) and the residual global risk (with the optimal controls applied). These values have been calculated by employing Algorithm 1.

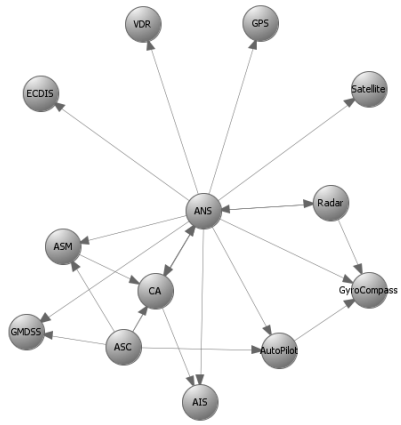


Figure 4. Autonomous ship—Navigational CPSs–$G_C(V, E)$–Control flow connections.

Each line of Table 5 represents one of the STRIDE threats. The first column represents the global initial risk (i.e., without any security controls in place) of the C-ES, as assessed by means of Algorithm 1. The second column represents each constituent CPS, and the third column the optimal set of security controls identified by means of Algorithm 2. Finally, the fourth column represents the residual risk (i.e., with the optimal set of security controls in place) of the C-ES, as assessed by applying again Algorithm 1 with the risks of each individual CPS updated according to the effectiveness of the applied controls.

Table 5. Optimal controls—Autonomous ship.

Threat	Initial Risk	Component	Controls	Residual Risk
Spoofing	1.651	ECDIS	Time Stamps (AU-8)	0.964
		ASM	Unsuccessful Logon Attempts (AC-7)	
		AIS	Remote Access (AC-17)	
		Radar	Security Assessments (CA-2)	
Tampering	1.615	AIS	Information Input Restrictions (SI-9)	1.087
		Radar	Tamper Protection (PE-3(5))	
		CA	Tamper Protection (PE-3(5))	
		ECDIS	Port and I/O Device Access (SC-41)	
		ASC	Tamper Protection (PE-3(5))	
Repudiation	1.555	Radar	Device Identification and Authentication (IA-3)	0.725
		AIS	Information System Component Inventory (CM-8 (4))	
Information Disclosure	1.629	AIS	Cryptographic Protection (SC-13)	0.89
		CA	Information System Component Inventory (CM-8 (4))	
		ECDIS	Protection of Information at Rest (SC-28)	
Denial of Service	1.373	AIS	Denial of Service Protection (SC-5)	0.89
		Radar	Fail-Safe Procedures (SI-17)	
		CA	Denial of Service Protection (SC-5)	
		ANS	Fail-Safe Procedures (SI-17)	
		ASC	Power Equipment and Cabling (PE-9)	
		ECDIS	Device Identification and Authentication (IA-3)	
		ASM	Fail-Safe Procedures (SI-17)	
Elevation of Privileges	1.129	ANS	Device Identification and Authentication (IA-3)	0.725
		AIS	Internal System Connections (CA-9)	
		ECDIS	Unsuccessful Logon Attempts (AC-7)	

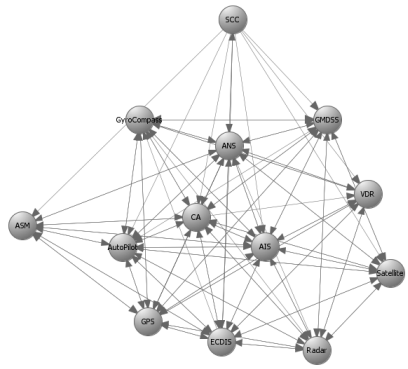


Figure 5. Remotely controlled ship—Navigational CPSs–$G'_I(V, E)$–Information flows.

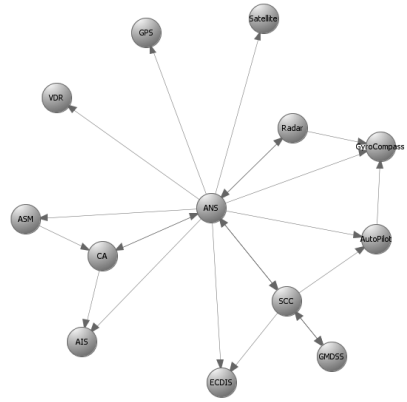


Figure 6. Remotely controlled ship—Navigational CPSs–$G'_C(V, E)$–Control flows.

6.3. Optimal Controls for the Remotely Controlled Ship

Remotely controlled vessels are equipped with CPSs that allow the control and operation of the vessel from the shore. Similarly with the autonomous vessel variant, the navigational CPSs of the remotely controlled ship are described by the directed graphs $G'_I(V, E)$ and $G'_C(V, E)$ in Figures 5 and 6. The SCC is a critical component in this variant of the C-ES, since the control and monitoring of the vessel critically depends on the SCC's normal operation. This is why the effect coefficients attain high values between systems that support the remote operations, such as the SCC, ANS, and ECDIS. All effect coefficients between the CPSs of the remotely controlled vessel are depicted in Table 6. Similarly to the case of the autonomous ship, the total effect coefficients have been calculated by means of Equation (13).

The security controls in the optimal set are selected from the initial list of available controls by applying the method described in Section 5. Table 7 depicts the optimal set of security controls per STRIDE threat and per CPS component. It also depicts the associated initial global risk (without controls) and the residual global risk (with the optimal controls applied). These values have been calculated in the same manner as the corresponding ones of the first C-ES variant.

Table 6. Effect coefficients—Remotely controlled ship.

	C-ES	AIS	ECDIS	VDR	ASM	RADAR	AP	C.A.	ANS	SCC
C-ES	0	0	0	0	0	0	0	0	0	0
ANS	0.208	0.208	0.208	0.208	0.208	0.106	0.208	0.208	0	0.208
SCC	0.75	0.5	0.75	0.5	0.5	0	0.75	0.5	0.75	0
ASM	0	0.071	0.071	0	0	0	0.071	0.321	0.071	0
AIS	0.041	0	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041
CA	0	0.295	0.045	0.045	0.045	0.045	0.045	0	0.295	0
ECDIS	0.05	0.05	0	0	0.05	0.05	0.05	0.05	0.05	0.05
RADAR	0.166	0.166	0.166	0.166	0	0	0.166	0.166	0.666	0.166
AP	0.045	0.045	0.045	0	0.045	0.045	0	0.045	0.045	0
VDR	0	0.062	0.062	0	0	0.062	0	0	0.062	0

Table 7. Optimal controls–Remotely controlled ship.

Threat	Initial Risk	Component	Controls	Residual Risk
Spoofing	1.952	SCC	Monitoring Physical Access (PE-6 (1))	1.663
		ASM	Unsuccessful Logon Attempts (AC-7)	
		AIS	Remote Access (AC-17)	
		Radar	Security Assessments (CA-2)	
Tampering	1.663	ECDIS	Device Identification and Authentication (IA-3)	1.04
		ANS	Port and I/O Device Access (SC-41)	
		Radar	Tamper Protection (PE-3(5))	
		CA	Tamper Protection (PE-3(5))	
		SCC	Physical Access Control (PE-3)	
		AIS	Information Input Validation (SI-10)	
Repudiation	1.828	AIS	Device Identification and Authentication (IA-3)	0.875
		Radar	Security Assessments (CA-2)	
		SCC	Non-repudiation (AU-10)	
Information Disclosure	1.828	AIS	Cryptographic Protection (SC-13)	1.47
		SCC	Information System Component Inventory (CM-8 (4))	
Denial of Service	1.622	ECDIS	Internal System Connections (CA-9)	0.99
		AIS	Information System Backup (CP-9 (1), (2), (3), (5))	
		CA	Denial of Service Protection (SC-5)	
		SCC	Denial of Service Protection (SC-5)	
		Radar	Security Assessments (CA-2)	
		ANS	Emergency Shutoff (PE-10)	
		ASM	Fail-Safe Procedures (SI-17)	
Elevation of Privileges	1.205	ANS	Device Identification and Authentication (IA-3)	0.875
		AIS	Internal System Connections (CA-9)	
		ECDIS	Unsuccessful Logon Attempts (AC-7)	

6.4. Discussion

The overall process followed to carry out the case studies is depicted graphically in Figure 7. In this figure, rectangles represent processing steps, and skewed rectangles represent input/output; solid lines link processing steps, whilst dashed ones link input/output to processing steps. The shaded area delineates the content of this paper.

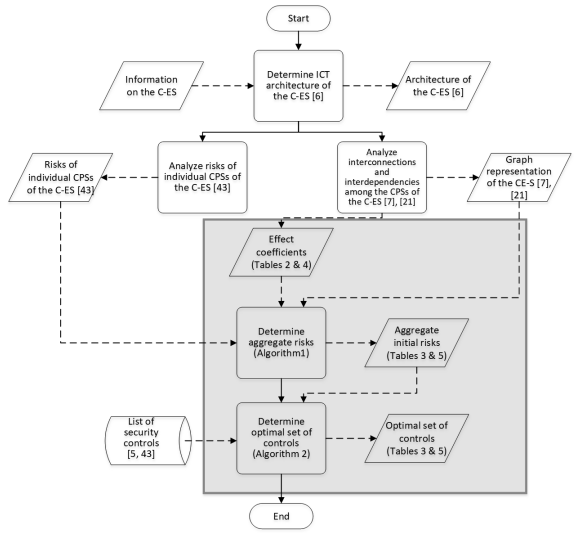


Figure 7. Overall process.

As can be seen in Table 5, in the case of the autonomous ship, twenty different security controls are recommended for application to seven of the ten navigational CPSs. The fact that these CPSs have been found in previous works [6,44] to be the most vulnerable onboard navigational systems, verifies the consistency of the proposed methods. Similarly, as can be seen in Table 7, twenty different security controls are recommended for application to six out of the ten navigational CPSs; again, these CPSs are the most vulnerable.

The optimal controls sets are different in the two variants of the C-ES. This reflects the difference in the level of autonomy of each variant: According to the IMO classification, the remotely controlled vessel lies at the second or third autonomy level, while the autonomous ship lies at the fourth level [52]. Different levels of autonomy mean different levels of interaction with humans and different levels of importance of the SCC in the ship's operation, which, in turn, mean different levels of risk for the same threat.

The security controls that are recommended by any automated decision support method, including the methods proposed herein, need to be *re-considered*, *consolidated*, and *checked for applicability* by domain experts and stakeholders together. The proposed methods enable the execution of what-if scenarios, including by modifying the initial list of the available security controls, and/or by modifying parameters of the genetic algorithm.

7. Conclusions

The growing utilization of highly interconnected CPSs in critical domains increases the attack surface, making the infrastructure more vulnerable to cyber attacks. In this paper, we model a complex CPS as a digraph in which nodes represent sub-CPSs and in which edges represent information and control flows among these subsystems. By leveraging this model, we proposed a novel method for assessing the aggregate cybersecurity risk of large scale, complex CPSs comprising interconnected and interdependent components, by using risk measures of its individual components and the information and control flows among these components. Building upon this method, we proposed a novel method, based on evolutionary programming, for selecting a set of effective and efficient cybersecurity controls among those in an established knowledge base, that reduces the aggregate residual risk, while at the same time minimizing the cost. We then used both methods to select optimal sets of cybersecurity controls for the navigational systems of two instances of the C-ES, namely the remotely controlled ship and the autonomous ship. These sets lead to

the definition of the cybersecurity architecture of such vessels. They have been found to be in line with previous results that identified the most vulnerable navigational CPSs of the C-ES, and to minimize the global residual risk. In the future, we intend to develop a software tool that will implement the proposed methods, and to use it to experientially examine the usability of the proposed approach with domain experts and stakeholders, in the C-ES and other critical application domains.

Author Contributions: Conceptualization, G.K., G.S. and S.K.; methodology, G.K., G.S. and S.K.; software, G.K. and G.S.; validation, G.K. and G.S.; formal analysis, G.K., G.S. and S.K.; investigation, G.K. and G.S.; resources, S.K.; writing—original draft preparation, G.K. and G.S.; writing—review and editing, S.K.; visualization, G.K. and S.K.; supervision, S.K.; project administration, S.K.; funding acquisition, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This paper has been partially funded by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773960 (DELTA project).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Giraldo, J.; Sarkar, E.; Cardenas, A.A.; Maniatakos, M.; Kantarcioglu, M. Security and privacy in cyber-physical systems: A survey of surveys. *IEEE Des. Test* **2017**, *34*, 7–17.
2. Cyber-Physical Systems Public Working Group (CPS PWG). *Framework for Cyber-Physical Systems*; NIST Special Publication 1500-201: Volume 1, Overview. Version 1.0; National Institute of Standards and Technology (NIST): Gaithersburg, MD, USA, 2017;
3. International Organization for Standardization, ISO. *ISO 31000:2018 Risk Management—Guidelines*; ISO: Geneva, Switzerland, 2018.
4. International Organization for Standardization, ISO. *ISO/IEC 27005:2018 Information Technology—Security Techniques—Information Security Risk Management*; ISO: Geneva, Switzerland, 2018.
5. Stouffer, K.; Pillitteri, V.; Marshall, A.; Hahn, A. Guide to industrial control systems (ICS) security. *NIST Spec. Publ.* **2015**, *800*, 247.
6. Kavallieratos, G.; Katsikas, S.; Gkioulos, V. Cyber-attacks against the autonomous ship. In *Proceedings of the SECPRE 2018, CyberICPS 2018*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11387, pp. 20–36.
7. Kavallieratos, G.; Katsikas, S.; Gkioulos, V. Modelling Shipping 4.0: A Reference Architecture for the Cyber-Enabled Ship. In *Proceedings of the Asian Conference on Intelligent Information and Database Systems*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 202–217.
8. Kouns, J.; Minoli, D. *Information Technology Risk Management in Enterprise Environments*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010.
9. Ali, S.; Balushi, T.; Nadir, Z.; Hussain, O. Risk Management for CPS Security. In *Cyber Security for Cyber Physical Systems*; Springer International Publishing AG: Cham, Switzerland, 2018; pp. 11–34.
10. Lyu, X.; Ding, Y.; Yang, S.H. Safety and security risk assessment in Cyber-Physical Systems. *IET Cyber-Phys. Syst. Theory Appl.* **2019**, *4*, 221–232. doi:10.1049/iet-cps.2018.5068.
11. Stellios, I.; Kotzanikolaou, P.; Psarakis, M.; Alcaraz, C.; Lopez, J. A Survey of IoT-Enabled Cyberattacks: Assessing Attack Paths to Critical Infrastructures and Services. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 3453–3495. doi:10.1109/COMST.2018.2855563.
12. Lamba, V.; Šimková, N.; Rossi, B. Recommendations for smart grid security risk management. *Cyber-Phys. Syst.* **2019**, *5*, 92–118. doi:10.1080/23335777.2019.1600035.
13. You, B.; Zhang, Y.; Cheng, L.C. Review on Cyber Security Risk Assessment and Evaluation and Their Approaches on Maritime Transportation. In Proceedings of the 30th Annual Conference of International Chinese Transportation Professionals Association, Houston, TX, USA, 19–21 May 2017; pp. 19–21.
14. Tam, K.; Jones, K. MaCRA: A model-based framework for maritime cyber-risk assessment. *WMU J. Marit. Aff.* **2019**, *18*, 129–163.
15. Tam, K.; Jones, K. Cyber-risk assessment for autonomous ships. In Proceedings of International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Glasgow, Scotland, UK, 11–12 June 2018; pp. 1–8.
16. Svilicic, B.; Kamahara, J.; Celic, J.; Bolmsten, J. Assessing ship cyber risks: A framework and case study of ECDIS security. *WMU J. Marit. Aff.* **2019**, *18*, 509–520.
17. Kim, Y.G.; Jeong, D.; Park, S.H.; Lim, J.; Baik, D.K. Modeling and simulation for security risk propagation in critical information systems. In *Proceedings of the International Conference on Computational and Information Science*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 858–868.
18. Kondakci, S. A new assessment and improvement model of risk propagation in information security. *Int. J. Inf. Comput. Secur.* **2007**, *1*, 341–366.
19. Feng, N.; Wang, H.J.; Li, M. A security risk analysis model for information systems: Causal relationships of risk factors and vulnerability propagation analysis. *Inf. Sci.* **2014**, *256*, 57–73.
20. Orojloo, H.; Azgomi, M.A. A method for evaluating the consequence propagation of security attacks in cyber–physical systems. *Future Gener. Comput. Syst.* **2017**, *67*, 57–71.

21. Wang, T.; Wei, X.; Huang, T.; Wang, J.; Valencia-Cabrera, L.; Fan, Z.; Pérez-Jiménez, M.J. Cascading failures analysis considering extreme virus propagation of cyber-physical systems in smart grids. *Complexity* **2019**, 7428458, https://doi.org/10.1155/2019/7428458.
22. Kavallieratos, G.; Katsikas, S. Attack Path Analysis for Cyber Physical Systems. In *Computer Security: ESORICS 2020 International Workshops, CyberICPS, SECPRE, and ADIoT, Guildford, UK, September 14–18, 2020, Revised Selected Papers*; Lecture Notes in Computer Science Book Series (LNCS), Volume 12501; Springer International Publishing: Cham, Switzerland, 2020; pp. 19–33.
23. König, S.; Rass, S.; Schauer, S.; Beck, A. Risk propagation analysis and visualization using percolation theory. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2016**, *7*, 1, http://dx.doi.org/10.14569/IJACSA.2016.070194.
24. Qu, Z.; Zhang, Y.; Qu, N.; Wang, L.; Li, Y.; Dong, Y. Method for quantitative estimation of the risk propagation threshold in electric power CPS based on seepage probability. *IEEE Access* **2018**, *6*, 68813–68823.
25. Zhu, B.; Deng, S.; Xu, Y.; Yuan, X.; Zhang, Z. Information security risk propagation model based on the SEIR infectious disease model for smart grid. *Information* **2019**, *10*, 323.
26. Malik, A.A.; Tosh, D.K. Quantitative Risk Modeling and Analysis for Large-Scale Cyber-Physical Systems. In Proceedings of the 29th International Conference on Computer Communications and Networks (ICCCN), Honolulu, HI, USA, 3–6 August 2020; pp. 1–6.
27. Theoharidou, M.; Kotzanikolaou, P.; Gritzalis, D. A multi-layer criticality assessment methodology based on interdependencies. *Comput. Secur.* **2010**, *29*, 643–658.
28. Theoharidou, M.; Kotzanikolaou, P.; Gritzalis, D. Risk assessment methodology for interdependent critical infrastructures. *Int. J. Risk Assess. Manag.* **2011**, *15*, 128–148.
29. Zhou, X.; Wang, F.; Ma, Y. An overview on energy internet. In Proceedings of the 2015 IEEE International Conference on Mechatronics and Automation (ICMA), Beijing, China, 2–5 August 2015; pp. 126–131. doi:10.1109/ICMA.2015.7237469.
30. Hong, Q.; Jianwei, T.; Zheng, T.; Wenhui, Q.; Chun, L.; Xi, L.; Hongyu, Z. An information security risk assessment algorithm based on risk propagation in energy internet. In Proceedings of the IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 26–28 November 2017; pp. 1–6.
31. Li, S.; Zhao, S.; Yuan, Y.; Sun, Q.; Zhang, K. Dynamic security risk evaluation via hybrid Bayesian risk graph in cyber-physical social systems. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 1133–1141.
32. Karbowski, A.; Malinowski, K. Two-Level System of on-Line Risk Assessment in the National Cyberspace. *IEEE Access* **2020**, *8*, 181404–181410.
33. Sawik, T. Selection of optimal countermeasure portfolio in IT security planning. *Decis. Support Syst.* **2013**, *55*, 156–164. doi:10.1016/j.dss.2013.01.001.
34. Viduto, V.; Maple, C.; Huang, W.; López-Peréz, D. A novel risk assessment and optimisation model for a multi-objective network security countermeasure selection problem. *Decis. Support Syst.* **2012**, *53*, 599–610. doi:10.1016/j.dss.2012.04.001.
35. Schilling, A.; Werners, B. Optimal selection of IT security safeguards from an existing knowledge base. *Eur. J. Oper. Res.* **2016**, *248*, 318–327. doi:10.1016/j.ejor.2015.06.048.
36. Baiardi, F.; Telmon, C.; Sgandurra, D. Hierarchical, model-based risk management of critical infrastructures. *Reliab. Eng. Syst. Saf.* **2009**, *94*, 1403–1415.
37. Gonzalez-Granadillo, G.; Garcia-Alfaro, J.; Alvarez, E.; El-Barbori, M.; Debar, H. Selecting optimal countermeasures for attacks against critical systems using the attack volume model and the RORI index. *Comput. Electr. Eng.* **2015**, *47*, 13–34.
38. Goldberg, D.E. *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed.; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1989.
39. Blickle, T.; Thiele, L. A Comparison of Selection Schemes Used in Evolutionary Algorithms. *Evol. Comput.* **1996**, *4*, 361–394. doi:10.1162/evco.1996.4.4.361.
40. Shostack, A. *Threat Modeling: Designing for Security*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
41. Zinsmaier, S.; Langweg, H.; Waldvogel, M. A Practical Approach to Stakeholder-driven Determination of Security Requirements based on the GDPR and Common Criteria. In Proceedings of the 6th International Conference on Information Systems Security and Privacy - Volume 1: ICISSP, 2020, Valletta, Malta, pp. 473–480. doi:10.5220/0008960604730480.
42. Kavallieratos, G.; Gkioulos, V.; Katsikas, S.K. Threat analysis in dynamic environments: The case of the smart home. In Proceedings of the 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), Santorini, Greece, 29–31 May 2019; pp. 234–240.
43. Seifert, D.; Reza, H. A security analysis of cyber-physical systems architecture for healthcare. *Computers* **2016**, *5*, 27.
44. Kavallieratos, G.; Katsikas, S. Managing Cyber Security Risks of the Cyber-Enabled Ship. *J. Mar. Sci. Eng.* **2020**, *8*, 768.
45. Microsoft. Chapter 3—Threat Modeling. 2010. Available online: https://docs.microsoft.com/en-us/previous-versions/msp-n-p/ff648644(v=pandp.10)?redirectedfrom=MSDN (accessed on 28-02-2021).
46. Rothlauf, F., Optimization Methods. In *Design of Modern Heuristics: Principles and Application*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 45–102. doi:10.1007/978-3-540-72962-4_3.
47. BIMCO; CLIA; ICS; INTERCARGO; INTERMANAGER; INTERTANKO; IUMI; OCIMF; WORLD SHIPPING COUNCIL. The Guidelines on Cyber Security Onboard Ships. Version 4. Available online: https://www.bimco.org/-/media/bimco/about-us-and-our-members/publications/ebooks/guidelines-on-cyber-security-onboard-ships-v4.ashx (accessed on 28-02-2021).
48. The President of the United States. National Maritime Cybersecurity Plan. White House Office, Washington D.C., USA, 2020. Available online: https://www.hsdl.org/?view&did=848704 (accessed on 28-02-2021).

49. Kavallieratos, G.; Diamantopoulou, V.; Katsikas, S. Shipping 4.0: Security requirements for the Cyber-Enabled Ship. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6617–6625.
50. Kavallieratos, G.; Katsikas, S.; Gkioulos, V. SafeSec Tropos: Joint security and safety requirements elicitation. *Comput. Stand. Interfaces* **2020**, *70*, 103429.
51. International Maritime Organization. Convention on the International Regulations for Preventing Collisions at Sea, 1972 (COLREGs). 1972. Available online: https://www.imo.org/en/About/Conventions/Pages/COLREG.aspx (accessed 24 January 2021).
52. International Maritime Organization. IMO Takes First Steps to Address Autonomous Ships. 2018. Available online: http://www.imo.org/en/mediacentre/pressbriefings/pages/08-msc-99-mass-scoping.aspx (accessed 21 September 2020).

ISBN 978-82-326-6790-1 (printed ver.)
ISBN 978-82-326-6979-0 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology