Applications

Kjeld Dittmann* and Mogens Blanke

Risk mitigation by design of autonomous maritime automation systems

Risikominderung durch das Design von autonomen maritimen Automatisierungssystemen

Design of autonomous maritime systems

Design von autonomen maritimen Systemen

https://doi.org/10.1515/auto-2021-0151
Received October 22, 2021; accepted March 18, 2022

Abstract: Autonomous marine surface vehicles rely on computer systems with computer intelligence making decisions to assist or replace the navigating officer. A fundamental requirement for the design and implementation of such a cyber-physical system is seamless, predictable, and secure interoperability between vendor-specific hardware and software subsystems. The article describes a system design that includes mechanisms to mitigate the risks and consequences of software defects, individual component malfunction, and harmful cyber interference. It addresses international regulations in the field and demonstrates a system design that can meet the requirements for safe behaviour in foreseeable events while also having the ability to call for human assistance if the autonomous system is unable to handle a situation. The paper presents a design for highly automated vessels with several inherent risk-reducing features, including the ability to isolate and encapsulate abnormal behaviours, built-in features to support resilience to unexpected events, and mechanisms for internal defence against cyber-attacks. The article shows how this is provided by a novel *middleware* that supports risk mitigation, dependability, and resilience.

Keywords: autonomy, architecture, safety, risk, distributed system, marine automation, modular design

Zusammenfassung: Autonome Überwasserfahrzeuge sind auf Computersysteme angewiesen, deren Computerintelligenz Entscheidungen trifft, die den Navigationsoffizier unterstützen oder ersetzen. Eine grundlegende Voraussetzung für den Entwurf und die Implementierung eines solchen cyber-physischen Systems ist die nahtlose, vorhersehbare und sichere Interoperabilität zwischen herstellerspezifischen Hardware- und Software-Subsystemen. Der Artikel beschreibt ein Systemdesign, das Mechanismen zur Minderung der Risiken und Folgen von Softwarefehlern, Fehlfunktionen einzelner Komponenten und schädlichen Cyber-Interferenzen umfasst. Er befasst sich mit den internationalen Vorschriften in diesem Bereich und zeigt ein Systemdesign, das die Anforderungen an ein sicheres Verhalten bei vorhersehbaren Ereignissen erfüllen kann und gleichzeitig die Möglichkeit bietet, menschliche Hilfe anzufordern, wenn das autonome System nicht in der Lage ist, eine Situation zu bewältigen. Im Beitrag wird ein Entwurf für hochautomatisierte Schiffe mit mehreren inhärenten risikomindernden Merkmalen vorgestellt, darunter die Fähigkeit, abnormale Verhaltensweisen zu isolieren und zu kapseln, eingebaute Merkmale zur Unterstützung der Resilienz bezüglich unerwarteter Ereignisse und Mechanismen zur internen Abwehr von Cyberangriffen. Der Artikel zeigt, wie dies durch eine neuartige Middleware ermöglicht wird, die Risikominderung, Zuverlässigkeit und Widerstandsfähigkeit unterstützt.

Schlagwörter: Autonomie, Architektur, Sicherheit, Risiko, verteilte System, Schiffsautomatisierung, modularer Aufbau

1 Introduction

Technological solutions supporting autonomous and unmanned crafts are gaining momentum in various domains, and the maritime domain is no exception. Autonomous marine crafts have been a research topic for

***Corresponding author: Kjeld Dittmann,** Automation and Control Group, Department of Electrical and Photonics Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark, e-mail: kjeditt@elektro.dtu.dk
Mogens Blanke, Automation and Control Group, Department of Electrical and Photonics Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark, e-mail: mb@elektro.dtu.dk

many years, with the MIT ARTEMIS being an early forerunner [43].

Marine autonomy research has been focused on algorithmic and technical developments, targeting autonomous craft in restricted areas where international rules and regulations were not a priority. This focus has changed. Today, the main technology drivers in the maritime industry are operational efficiency and safety [21] and the introduction of autonomous technology onboard ships have the potential to enhance both. Under consideration is a temporarily unattended bridge, which would allow the navigating officer to attend to other duties while being alerted when human attention is required. An extension of this is the possibility of having the human proxy located at an Remote Control Centre (RCC). Trials with offshore vessels and tugs have shown how the duty of the bridge crew can be temporarily substituted by a manned RCC.

One of the consequences associated with the introduction of autonomous systems is complexity, both in the system design and implementation. Therefore, the risk factors change towards software development and validation [20, 51] as well as the cyber-physical dimensions [48, 54]. The development of system-wide design principles and tools to meet these challenges is only at an early stage. Furthermore, introducing new technologies in a highly safety-regulated domain imposes other challenges, which need to be carefully considered.

This paper is focused on a systematic design that mitigates risks by design and considers the regulatory regime that is in place to ensure safety. We address the measures that are needed to mitigate and provide resilience to the risks associated with software defects, malfunctions, and harmful cyber-interference. We discuss the implications of international regulations in the field and show a path towards a design that can meet the requirements for safe behaviour. We develop a design paradigm for autonomous systems that can isolate and encapsulate not normal behaviour and have inherent features that make it resilient to rarely occurring events and provide internal defence against cyber-attacks.

The remainder of this article is structured as follows: Section 2 provides the context and introduces the target vessel for this research, i. e., an autonomous harbour bus. Section 3 introduces the aspects of safety assessment. In Section 4 the changing risk landscape is addressed and the derived consequences for the design process are discussed. A novel software architecture, ensuring system-level resilience, is introduced in Section 5. Finally, Section 6 offers the conclusions.

Table 1: List of Acronyms.

Notation	Description
ACS	Autonomous Coordination Supervisor
AIS	Automatic Identification System
ANS	Autonomous Navigation Supervisor
APS	Autonomous Platform Supervisor
DHS	Distress Handling Service
ENC	Electronic Navigational Chart
GNSS	Global Navigation Satellite System
IACS	International Association of Classification Societies
IMO	International Maritime Organization
MASS	Maritime Autonomous Surface Ship
RCC	Remote Control Centre
SAS	Situation Awareness Service
SCC	Ship Control Centre
SFU	Sensor Fusion
SHP	Short Horizon Planner
STCW	Seafarers Training Certification and Watch-keeping
UN	United Nations
UNCLOS	UN Convention of Law of the Sea
VCS	Voyage Control System

2 Autonomy system in context

The business case for introducing maritime autonomy depends on various elements, e. g., ship type, size, and operational profile, where some of them cannot be quantified at present. The timing might therefore be very difficult to foresee, although some predictions indicate that the first generation of autonomous vessels will be in operation by the end of 2021 [6]. However, if the technology provides a solution that has a lower total cost of ownership, equal safety level and can be approved by the authorities – it is likely that autonomous solutions will partly complement and partly substitute traditional ships if the overall risk can be reduced.

Terms and definitions within the domain of maritime autonomous surface ships are still to be settled and, in this article, the term autonomous vessel does not equal an unmanned vessel. Equally, a floating construction with its own propulsion and steering system serves as a definition of a vessel or ship and when adding self-governing capabilities, it is categorised as *autonomous*.

2.1 Regulatory regime

International legislation is governing the co-existence and operation of vessels at sea. Vessels are only granted permission to operate if they comply with international regulations [53]. The International Maritime Organization

(IMO) is the agency of the United Nations (UN) that is responsible for the maritime domain. The international law of the sea is anchored in the United Nations Convention of Law of the Sea (UNCLOS), and IMO has the mandate to create a fair, i. e., no favourable treatment, and an effective regulatory framework, i. e., can be complied with and being policed without stipulating technical solutions. Considering that the purpose of the autonomous system is to replace or supplement the navigating crew of a vessel, it is important to notice that the IMO regulations depend on and utilise the framework of skilled seafarers. This is a framework that safeguards compliance of UNCLOS and IMO conventions and represents the vessel owner and the flag state, i. e., the UN member state. IMO regulations address the operational level as well as the management level and divide the needed competencies to undertake the safe operation into three groups, i. e., operation and controlling the vessel, navigating the vessel, and handling the cargo. The onboard autonomous system does not necessarily need to have the capabilities of both operation and management, i. e., to assume responsibility as the watch-keeping officer will only require the operational competencies, however, the combined system needs to cover both. It is foreseen that the competencies related to the management level can be fulfilled with a backup system at an RCC.

The concept of periodically Unattended Machinery Space is well established and it covers design arrangements, monitoring capabilities, situation assessment and an alert management system that enables the engine crew to leave the engine space and/or the engine control room and have near-normal working hours. The emerging term periodically Unattended Bridge Space is targeting an equivalent solution for the bridge crew, which must provide the same degree of safety as the watch-keeping personnel it replaces to get acceptance. Chapter 2 of IMO's Seafarers Training Certification and Watchkeeping (STCW) publication [32], Table A-II/I from page 32, specifies the competencies required to operate as a navigational officer.

The current rules and regulations do not support the deployment of an autonomous vessel and IMO has therefore assessed the degree to which the existing regulatory framework might be affected to address Maritime Autonomous Surface Ships (MASS) operations [38]. In support of the development of maritime autonomous systems, IMO has published guidelines for conducting trials [36]. IMO's Strategic Plan [35] provides directions for the, *"Integrate new and advancing technologies in the regulatory framework"* [37] that in combination with [30] introduces a modular design, certification standards & rules.

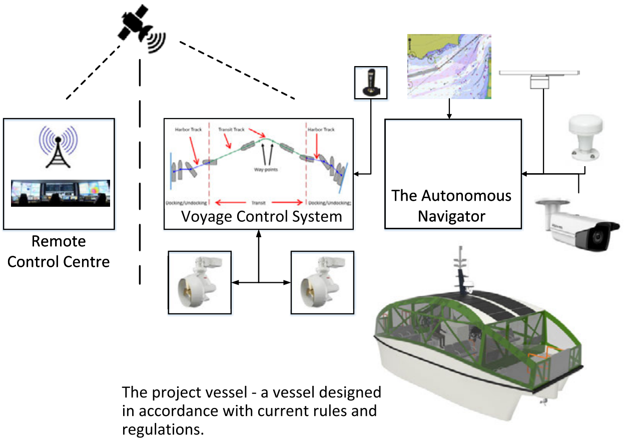


Figure 1: The reference vessel with the additional components required to support autonomous operation – project vessel simplified.

When introducing large and highly complex automation systems onboard, the prescriptive approval regime is being changed to a goal-based regime where risk assessment is essential for achieving an approval [52]. The safety verification of a product using a goal and risk-based approach, [33], will be demanding, but necessary to certify that the technology and its implementation is safe [7]. Critical for the introduction and acceptance of new technology solutions, is to identify a stringent path for how the transition and the expansion of highly automated vessel, i. e., dealing with known processes and actions, to vessel having computer intelligence dealing with autonomy, i. e., dealing with the unknown.

2.2 ShippingLab GreenHopper

A conventional vessel design, that is, a vessel designed and certified by the flag state in accordance with current rules and regulations, serves as the baseline for the project vessel.

The method of expressing the path to follow [28] and the performance & test standards for control systems [27] was adopted from conventional IMO vessels and guides the allocation of functions into vessel control, virtual captain, and shore-side supervision.

Figure 1 illustrates the reference vessel with the additional components required to support autonomous operation. The additional main components are:

Remote Control Centre. During periods of unattended bridge operation of the vessel, the shore-based crew of an RCC will undertake the role of the Watch Keeping Officer.

Table 2: GreenHopper Key Parameters.

GreenHopper Key Parameters
Length: 12.2 m
PAX capacity: 25
Wheelchair capacity: 1
Bicycles capacity: 4
Baby carriage: 1
Crew members (space for 2)
Propulsion: 2 electric hydraulic rotate-able thrusters
Energy storage: Batteries
Speed (@MCR): 8 knots
Hull: Catamaran hull in a double-ended design
Displacement: 15 t

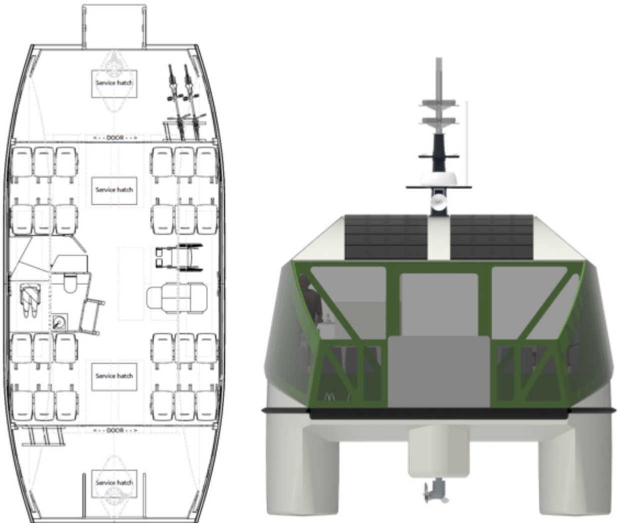


Figure 2: Rendering of the ShippingLab GreenHopper.

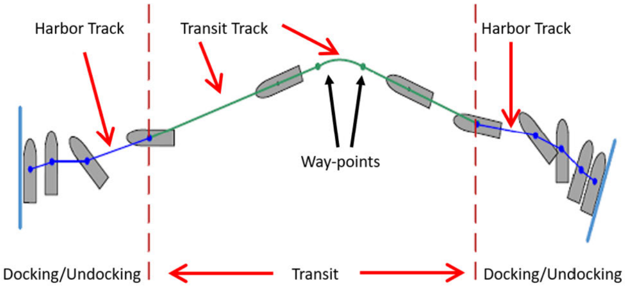


Figure 3: The route to follow.

Communication Link. The situation awareness requirement originating from the RCC puts stringent reliability and capacity requirements on the communication system.

Voyage Control System (VCS). Two steerable thrusters are used by the Voyage Control System (VCS) to control the movement of the vessel. The system has been equipped with joystick control to ease thrust allocation in manual mode.

Autonomous Navigator. Safe navigation is undertaken by the autonomous navigator in periods of unattended bridge operation. The system uses a 360-degree camera (daylight and infrared), radar (X and W bands), and lidar coverage for object detection, and cross-validates the detections using information from the Electronic Navigational Chart (ENC), Global Navigation Satellite System (GNSS), and Automatic Identification System (AIS).

The development of systems and modules described throughout this article will conclude with the installation in the first quarter of 2022 and trials will be conducted during the month of August 2022. The ShippingLab GreenHopper, i. e., the project vessel, will go into service in autumn 2022, however, with an onboard human on the loop. The key features of the project vessel are shown in Fig. 2 and Table 2.

2.3 Voyage control system

Trials with highly automated vessels have demonstrated that the onboard control system can steer a vessel from the quay in one port and to berth it in another [3, 49]. The autonomous system utilises a sub-system, the vessel control system, VCS, that provides closed-loop control and allocates forces and moments from available actuators on the vessel. The VCS provides a consistent and repeatable automated execution of an optimised vessel trajectory, defined by planned path and speed.

Figure 3 shows how a route to follow is defined by a set of way-points [28] connected by line segments. The VCS follows this route using a track control concept. The VCS provides an automated vessel manoeuvring function that performs the automated vessel transit, docking, and undocking functions. The route can be amended by the autonomous system when an evasive manoeuvre is needed to minimise the risk of collision. The route description used for the track outside the harbour area is equivalent to the route information used by a track control system as defined in [27]. This part of the route is referred to as the transit track in Fig. 3. The manoeuvring inside the harbour area and the final berthing track is performed by a control system equivalent to a slow speed track follower used in dynamic positioning systems. The way-points describing harbour and berthing tracks have been extended with additional parameters, e. g., the centre of rotation. The VCS has four modes: automatic, joystick/tiller/coordinated, manual/individual mode, and fail-safe. The automatic mode has three sub-modes, i. e., transit, harbour, and berthing. The joystick mode has two sub-modes, i. e., Ship Control Centre (SCC) and RCC control. In manual mode, the vessel is handled by the crew on board, while in automatic mode, the autonomous onboard system provides the track to follow, i. e., the VCS is controlling the vessel. In remote mode, the vessel is handled by the RCC.

In this article, the VCS is a generic term encompassing the following functions:

Heading control is a function that keeps a ship's heading at the pre-set heading. Where the heading is the direction in which the longitudinal axis of the ship is pointed, defined by the angle between the meridian through its position and the fore-and-aft line of the ship, expressed in angular units from true north. The function is often referred to as an autopilot [39].

Track control is a function that maintains control of the ship's movement along a track, corrections made by the controller to compensate for wind, drift and other influences are based on the cross-track error and not only on the bearing to the destination waypoint. The track control system is active at a higher speed, where direct sway control is not possible. The vessel's heading and lateral position relative to the track are obtained by using rudders or steerable thrusters [27].

Auto Berthing/Auto Un-Berthing is a function that maintains control of the ship's movement along a track inside the harbour and related areas as well as the berthing process. Full three-axis position and heading control is possible once the vessel speed is reduced and achieved by allocating thrust to all available thrusters, i. e., a slow speed track follower function [18].

As for a conventional vessel, the pre-departure voyage planning for the autonomous vessel will be performed, taking parameters such as weather and the desired arrival time into account. IMO-compliant voyage planning tools [29] are well established in the maritime industry and commercially available.

2.4 Remote control centre

There are numerous reasons why an RCC will need to be part of an overall design for the introduction of autonomous vessels. As elaborated in section 4.3, the liability issue will require a human proxy to be ultimately responsible. Another important factor is the shared and sliding responsibility, as highlighted in [17], which describes the mapping from STCW competence requirements to the functionality of the autonomy modules. Assuming that the competencies related to the management level of STCW [32] will be fulfilled by an RCC, or as a backup for the operational competencies, a co-design of the RCC and the autonomy system onboard the vessel must be performed.

The RCC requires the following information in order to provide the needed situation awareness for the operator, [21, 44, 47]:

Voyage information. The ENC shows overlay, waypoints, course to steer and the allowed cross-track corridor (voyage plan – task 1(a) in Table A-II/I [32]).

Navigational information. This includes position, course, speed, and heading. Route-related information, e. g., time and distance to the next waypoint, and weather services (conduct the voyage – task 1(b) in Table A-II/I [32]).

Object detection. Information from the Autonomous Navigation Supervisor (ANS) identifies other ships, their anticipated course and speed, and any risks (maintain safe navigational watch – task 2 in Table A-II/I [32]).

Other parameters that are important to be transferred from the ship to the RCC are related to the Autonomous Platform Supervisor (APS): 1.) Dynamic information (ship motions), 2.) Safety and emergency, 3.) Propulsion system status, and 4.) Condition of cargo and vessel's intact stability [2, 31].

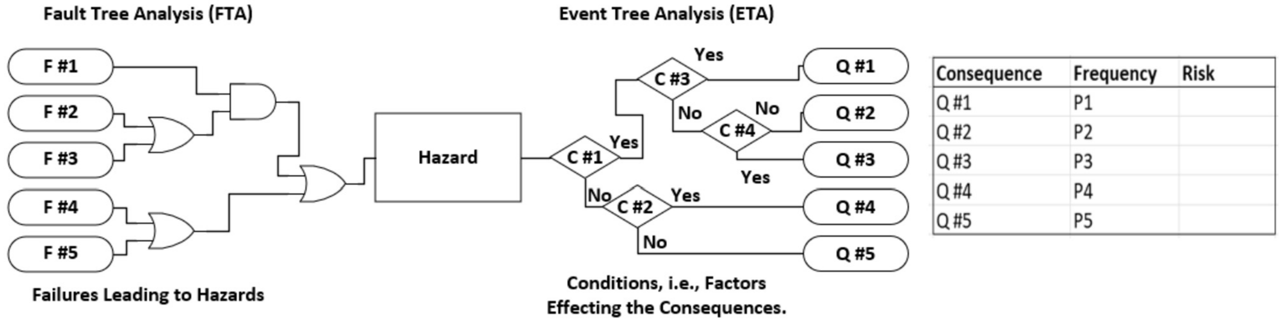
The interaction with the onboard autonomous supervisor will be detailed in Section 5.

2.5 External connectivity

The creation of a robust infrastructure for autonomous vessel operation, will in addition to the vessel-based and the land-based infrastructure, require a communication infrastructure. Typical hazards related to external communication are listed in [14] along with design guidelines, typical capacity requirements, and latency expectations.

3 Safety assessment

The proposed approach aims at providing a framework by which the safety of a ship, which utilises an autonomous system and is operating without or with a reduced crew, can be evaluated with respect to risk. The recommenda-


Fault Tree Analysis (FTA)
F #1
F #2
F #3
F #4
F #5
Failures Leading to Hazards
Hazard
Event Tree Analysis (ETA)
C #1
C #2
C #3
C #4
Yes
No
Q #1
Q #2
Q #3
Q #4
Q #5
Conditions, i.e., Factors
Effecting the Consequences.

Consequence	Frequency	Risk
Q #1	P1	
Q #2	P2	
Q #3	P3	
Q #4	P4	
Q #5	P5	

Figure 4: Generic Risk Assessment Methodology. Faults are denoted F, conditions are C, consequences are Q, and Frequencies P.

tions are based on investigations of currently available safety assessment techniques, existing class rules and guidelines [7, 14, 33, 36] for specialised installations.

The introduction of autonomy will significantly change the operational philosophy of the vessel, both in terms of the automation and computer intelligence employed, as well as the manning philosophy. The impact on vessel safety caused by the changed operational regime, along with the reliability of the systems provided to facilitate the autonomy vessel, imposes a risk pattern that needs to be addressed. It is suggested that the safety assessment of an autonomous vessel is being sorted into design & construction and operational aspects. The breakdown allows differentiating between aspects that are largely unchanged from a conventional ship and those that are either added or significantly different. The first category covers aspects of mechanical integrity, and the second category concerns the safe operation of the vessel.

Evaluation of hazards and risks are essential elements in a safety assessment. A hazard is *"A condition with a potential for human injury, damage to property, damage to the environment or some combination of these"*, and risk is defined as *"The likelihood of a specified undesired event occurring within a specified period or in specified circumstances (frequency or probability)"* [40].

The basis for hazards and risk assessment is a definition of operation and system structure.

3.1 Operation and system structure

For conventional vessels, prescriptive rules exist for the design. Certification is obtained when design rules are adhered to, and a risk assessment is rarely needed. The safety assessment for an autonomous vessel is different. Prescriptive rules do not exist, and software complexity implies that traditional testing is supplemented by a risk assessment.

For the autonomous vessel, the operational regime changes. Computer intelligence supplements or replaces traditional roles of humans on board [32] and human supervision is moved to an RCC. Furthermore, cyber vulnerability is an undesired side effect.

The change in operational conditions is primarily related to the remote operation. Training & education of the RCC personnel and their capabilities need to be equivalent to or exceed those for onboard personnel [32]. The analysis of the appropriate manning level of the vessel and the shore side to ensure safe operation during normal operation, as well as foreseeable emergencies, is outside the scope of this article. However, the manning and safety analyses are interlinked, and the overall operational safety assessment depends on an alignment between the two [21, 47].

Regarding system structure, the first critical question to seek answers to is related to reliability and whether or not the applied systems are adequate to perform the required functions, i. e., is it fit for purpose [42]. The second critical question concerns the situation where the functions of the autonomy system might fail. A safety assessment must show whether or not safety measures are in place to avoid critical consequences and escalation.

3.2 Hazards and failure modes

Hazards are events that will have adverse effects on safety and performance if they are not mitigated. Figure 4 illustrates the assessment method applied when creating lists of external and internal hazards that must be identified as part of the required assessments of risks. Among others, the classification society Bureau Veritas has issued a guideline for risk and technology assessment, see Section 2 of [14].

A "Hazards and Operability" analysis is a standard procedure to assess an automated process. It identifies

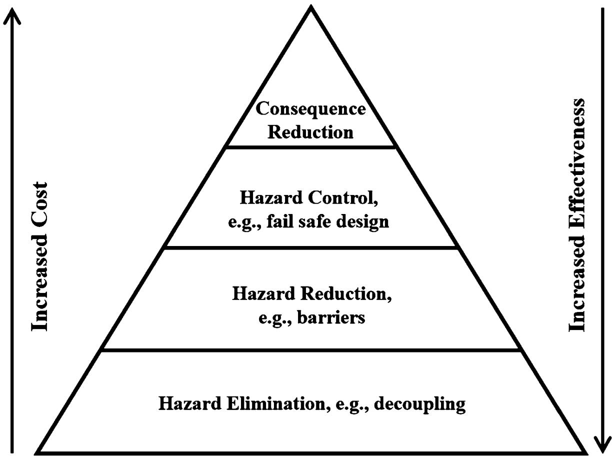

Consequence Reduction
Hazard Control, e.g., fail safe design
Hazard Reduction, e.g., barriers
Hazard Elimination, e.g., decoupling
Increased Cost
Increased Effectiveness

Figure 5: Cost and Hazard Pyramid.

where risks need to be mitigated, for the hazards that have been identified. Hazards analysis operates at an upper level of functionality and events. At the more detailed level, Failure Modes, Effects, and Criticality Analysis (FMECA) digs deeper and shows how the failure of individual components ripples through a system and describes its effects. The evaluator assesses the criticality of the events and the design team changes design details with the result that critical events can be mitigated.

The coverage of hazards and failure modes is an uncertain factor in the assessment as it relies on the experience of the assessor. To overcome the issue of coverage, stringent analysis of behaviours and system topology was introduced, with solution tools based on mathematical graph theory.

3.3 Analysis of behaviours

The *structural analysis* methodology guarantees inherently correct analysis given that normal behaviours are specified at a sub-system level, supplemented with the topology of signal flow between connected sub-systems [9]. Based on a formal description of the normal behaviours of components in a system, it offers graph-based techniques to analyse violations of normal behaviours at a sub-system and component level. With a complete description of the normal behaviours in a system, and the graph that describes the relations between components, the structural analysis provides distinct answers related to the detailed design. The method identifies which violations of normal behaviour can be detected, isolated to an individual component, and mitigated [11]. It shows where functions need to be fail-operational, e. g., by redundancy, to mitigate undesired effects.

Furthermore, the graph-theory based tools for structural analysis generate sets of *residuals*, which can be used to diagnose whether or not constraints are violated. This feature is used to make sensor fusion resilient to sensor defects, subsystem defects, and cyber-attacks that violate normal system behaviour [45].

Fault-tolerant sensor fusion is used to mitigate defects, e. g., to avoid using a sensor signal that has temporal or permanent defects in the information it provides. Solutions in the maritime field based on this methodology include fault-tolerant station-keeping control [8], risk mitigation against mooring line breakage [23, 24] and total system analysis for offshore position mooring [10]. Said references showed the structural analysis approach to obtain detection and mitigation of vessel-related phenomena. The analysis was extended in [45] to include sea chart information, surrounding objects at sea and on land using radar and camera detection. This enabled detection of information defects in a wider sense by fault-tolerant sensor fusion, also in cases where the reason for defects in information was caused by a cyber-attack. These design principles are being used for autonomous navigation, both for resilient sensor fusion and to make the autonomous system aware of the status of all sensor systems.

Structural analysis is a tool with binary decisions. Given the desired probabilities of false alarm and missed detection, a hypothesis test decides if a constraint is violated. If it is violated, the associated component is considered defect. Risk assessment techniques must be applied when the probability of failure of sub-systems is required and the end effects need to be classified by likelihood and severity.

3.4 Risk assessment and uncertainty

Estimating the risk and assigning criteria for the risk assessment will be predisposed to uncertainty. The challenge in computing risk originates from three main elements: the inaccuracy in estimating the likelihood of a hazard, the exclusion of scenarios, and the impact of the consequences. The identification of accident scenarios for new solutions and technologies is highly dependent on the composition and quality of the team that performs the analysis when introducing autonomy at sea. Difficulties include the shared responsibility between computer intelligence and a master mariner, as well as navigation in coexistence with conventional vessels. The contribution of this to the overall risk will emerge over time as data accumulates.

The consequence assessment will equally be prone to uncertainty due to the inaccuracies in the assumptions made. Full-scale tests are performed in some industry domains in order to improve consequence assessments [41]. However, the scenario complexity in the maritime domain causes the cost of such tests to be exorbitant, in particular when introducing new technologies, but also due to the small number of identical vessels, i. e., sister ships, being manufactured per year.

The availability of failure data that accurately represents the system in a setting equivalent to the one being assessed is hence a major difficulty in estimating the likelihood of an accidental event. New procedures are therefore needed for risk assessment of autonomous vessels in a changing landscape of risk.

4 Elements of a changed risk landscape

The changing risk landscape originating from the introduction of autonomy functions requires that the traditional risk calculation be supplemented with measures and processes that do not require accumulated accident data. Obtaining a high degree of safety and security in highly complex computer-based systems must be an integral part of the design process, equal to applying quality control measures [41]. Certification and approval of computer-based products and solutions installed onboard conventional vessels are assessed by prescriptive rules and procedures, i. e., performance and test standards. Nearly all the applied test standards originate from electromechanical systems assessment techniques that are not applicable to complex computer-based systems. As a result, many accidents are related to design flaws in the software, although they are not classified as such.

4.1 Cyber security

The landscape of cybersecurity-related safety aspects in maritime transport is changing from unlikely to occur with regularity. The protection methods that are currently being applied, i. e., solutions utilising firewalls and virtual private networks, have shown to be insufficient in numerous cases [4].

Two significant factors have progressively changed the effect of utilising computer-based solutions. The first is related to the onboard systems and the increasing connection of homogeneous automation systems by computer networks, i. e., systems utilising similar operating systems and programming languages. Traditional marine automation is vendor-specific and therefore non-homogeneous. Such intrinsic incompatibility protects the non-homogeneous system from computer viruses spreading, and hence limits the risk of a total vessel breakdown. Likewise, the one-off design also makes it robust against mainstream attack tools [16, 25]. The second issue is related to the introduction of external communication that exposes the information technology as well as operational technology of the vessel to cyber threats from attackers that do not need to pass through the physical barrier of the vessel. It has turned the connected vessel into a subject of interest by cyber-criminal organisations. The development in cyber security is of great concern and has ignited comprehensive work within governments, international organisations and agencies [5, 13, 19, 34, 46].

4.2 Human factors

There is an increased complexity in the interactions between automation systems and humans, i. e., the crew is increasingly sharing control of the vessel with automation systems, and the borderline of control responsibility might even be shifting depending on the situation. These changes are leading to new types of risks, such as mode confusion and automation “surprises” [21]. The annual report from the European Maritime Safety Agency [22] indicates that only a minor part of the incidents are related to equipment and component failures, while up to two-thirds of accidents have been classified as caused by human error. However, many of these accidents could more accurately be classified as originating from the working conditions of the crew [21], the situational awareness, and their interaction with computer intelligence. Human error is an indication of a problem deeper inside the system. Predictability, transparency, and intuitive reactions are essential elements of system design.

Minimising the risk related to the interactions between automation systems and humans is an extensive task. Section 5.3 addresses the functional mapping onto a control hierarchy [12] and utilising the principles of the STCW, see [32].

4.3 Liability

It is argued [15, 50, 55] that autonomous vessels operating in international trade are likely to have the same liability and navigational rights as conventional manned vessels if they follow the rules and regulations that apply to conventional manned vessels [33, 53]. It is further argued that the shipbuilder and component manufacturers are liable if the accident and related loss can be justified. The matter of concern is to be determined before a MASS will be able to operate. While this is considered outside the scope of this article, it is duly noted by the authors, and the underlying assumption is that there will be a human proxy, likely located at an RCC, which is held accountable.

5 A sovereign based vessel operating system

This section introduces a *middleware* layer that is in control of all network traffic and message communication between modules in the autonomous system. The purpose of the middleware is to secure the reliable and safe operation of the vessel in all conditions, including unintended issues caused by onboard software or cyber events. The principle used is to have an architecture of sovereign agents (modules). They have predefined roles and strict control of which inputs and outputs they handle.

5.1 Barriers and layer of defence

The outer shells of defence provided by the sovereign architecture of modules have the purpose of securing that no traffic to or from a module is possible unless it has predefined senders/destinations. The format of messages is defined by a module specification. When complementing modules with safety critical services with model-based diagnosis and fault-handling mechanisms, the integrity can be maintained, even in the situation where outer defence lines collapse. Model-based detection and fault-handling mechanisms are extensively addressed in the literature on diagnosis and fault-tolerant control [9]. The principles have been applied and demonstrated in aerospace [1] and aviation [26]. Irrespective of the source of faults or failures, e. g., software-based functions, instruments, or hardware, caused by external events or cyber-attacks, the autonomous vessel needs to be resilient. Autonomous systems have, in addition to the methods based on models of the physical system, the option to utilise cross validation of information from intelligent sensors that monitor the surroundings. These include object detection with cameras in the infrared and visible parts of the spectrum, and radars in the cm and mm range.

The fault-tolerant sensor fusion utilises a 360-degree camera and radar coverage, the ENC, and the AIS to cross validate and exclude a suspect sensor before safety and dependability become at risk. The fault-tolerant sensor fusion uses buoy and beacon positions from the ENC and land contours obtained from radars in near coastal situations to eliminate spoofed GNSS sensor information and manipulated AIS messages. Utilising the allocation of functions from the project vessel as an example, Fig. 7, the ANS module evaluates the identified objects within the awareness zone around the vessel. This information is provided to the Situation Awareness Service (SAS) in the form of a dynamic list of detected and classified objects or targets. The tracking of objects is performed by the Sensor Fusion (SFU) module, utilising radar, AIS, camera-based object detection, and ENC. The SAS performs understanding and anticipation in sub-modules. The semantics of the current situation are supplemented by the understanding module.

5.2 Allocation of functions in a control hierarchy

Establishing a generally accepted structure or architecture supplemented with interface standards is fundamental in managing abstraction and the foundation for risk mitigation, i. e., creating a framework for where the individual sub-functions belong and the allocation of functions in terms of a control hierarchy, [12]. Figure 6 illustrates how to deal with the increased complexity, i. e., by locating the functions and aggregating the information. A rigid modular design framework strengthens the test strategy of module & regression tests, and it enables automated consistency & interoperability verification. The modular design is an important risk-reducing factor in support of the incremental integration and certification process onboard the vessel. The control hierarchy has been successfully applied in space systems, where the operator is in the loop when ground station passage takes place. For agricultural robots, a human can be in the loop from a remote control centre if required. For ships, a human in the loop is the Officer on Watch present on the bridge.

5.3 Mapping of the autonomy system

In Fig. 7, the module has been schematically assigned to the layers of the control hierarchy, and the Table 3 details the control relations between the modules.

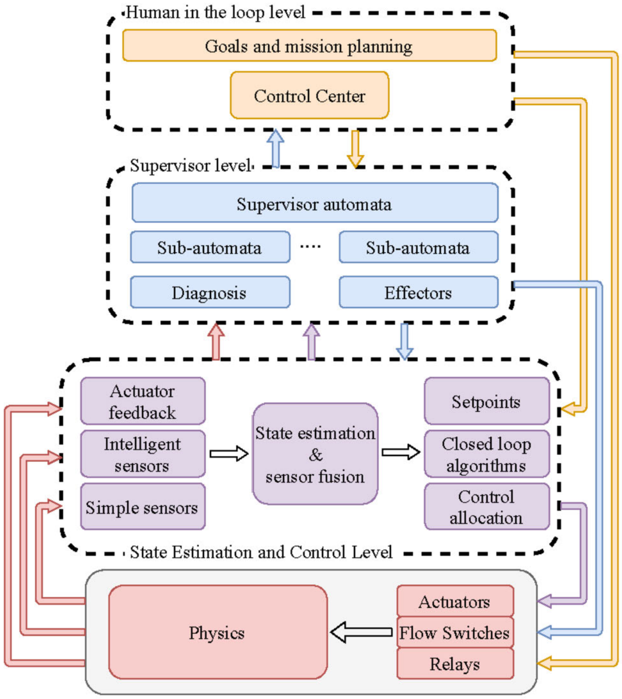


Figure 6: The Control Hierarchy for a system with supervision.

The notation in Table 3 is interpreted as follows:

$$A \leftarrow B : A \text{ receives input from } B$$

$$A \leftarrow \{B, C\} : A \text{ receives input from } B\&C$$

$$A \leftarrow (SCC \vee RCC) : \text{human proxy on the loop}$$

The functionality of supervision in the autonomous system replicates the established procedures required in human watchkeeping. The mapping from IMO regulations to autonomous functionalities was the subject of [17].

5.4 System integrity and interoperability

Alternative fuels are often highlighted as the path to the green transformation of the maritime industry. However, equally important is the utilisation of the limited availability of these fuels in a near-optimal way. An autonomous vessel can be classified as a system of systems that need to cooperate, i. e., share information in a normalised infrastructure. To gain the optimising benefits of the cyber-physical system, it needs to provide interoperability between domain-specific hardware and software in a seamless, predictable, and secure way. A comprehensive assessment of existing frameworks like ZeroMQ, RabbitMQ and ROS revealed a high level of complexity and dependency on a central message exchange service. This approach does not meet our system criteria for an effective and robust integration framework.

The generic mechanisms to ensure safety in control structures were addressed in the design guide by [41] for spacecraft systems.

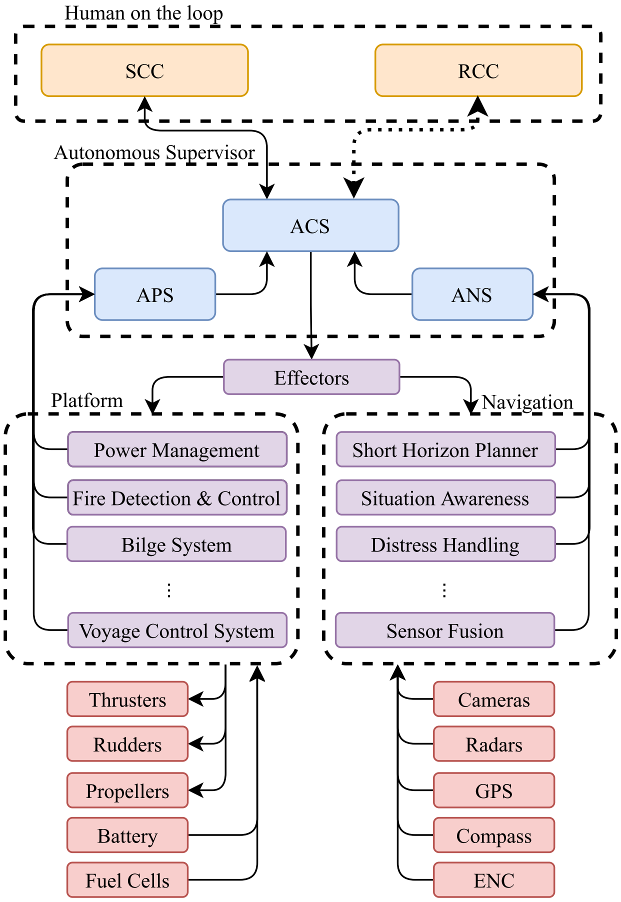


Figure 7: Allocation of modules in the Control Hierarchy. Autonomous supervision for overall coordination, navigation, and platform control with connections to modules that execute core functions for safe navigation and platform control. Outside vessel communication is indicated by a dotted line to the RCC.

Table 3: Control flow for autonomous supervision.

Control flow
ACS ← ANS ← {SHP, SAS}
ACS ← ANS ← {SAS, SFU}
ANS ← SFU
ANS ← {SAS, SFU}
ACS ← (SCC ∨ RCC) ← ACS ← APS
ACS ← (SCC ∨ RCC) ← ACS ← ANS ← DHS
(SCC ∨ RCC) ← ACS
ACS ← (SCC ∨ RCC) ← ACS ← ANS ← DHS
VCS ← ACS ← ANS ← {SAS, SHP}

Table 4: Key Design Features of the middleware in support of system integrity and cyber resilience.

Feature	Middleware Key Features
1	Worst-case performance design
2	No “master” component
3	Segmentation
4	Redundancy
5	End–to–end message encryption
6	System wide accurate time synchronisation (1 ms)
7	Connection oriented protocol
8	Immediate acknowledge
9	System integrity monitoring
10	Connection to external systems via white-list mapping
11	Network segment re-routing within 5 ms
12	All standard service ports are close
13	Only entry via the Middle-Ware message parser
14	Manage the topic updates and flag in case of malfunction
15	Manage the active-running and hot application topic update

For marine systems, we adhere to these guides where applicable, but in addition, we have an integration mechanism that is operating system and programming language agnostic. A novel *middleware* has been designed as a generic technology for autonomous vessels. It is based on a publish/subscribe communication protocol. The middleware ensures that modules do not impose constraints on the configuration and communication between other modules in the system. This implies, as an example, that, although the understanding and anticipation sub-modules are part of the same overarching module, the sub-modules do not need to be located on the same physical hardware node. Apart from improved reliability, this design also enables good testability as modules can be tested in isolation, without the entire system being available. The various design features are listed in Table 4.

The need for reducing the surface of attack was highlighted in section 4.1. The middleware has therefore a rigid policy for joining the networked autonomy system that is supported by features 10, 12, and 13. Asset mobility [46] has been addressed by features 4 and 15, enabling floating hosting of services, while features 2, 3, and 5 address the more general cyber-related risks. Features 7, 8, 9, and 14 have been implemented in support of system integrity, and the features 1, 6, and 11 were implemented to support the applications.

The control structure mapping is listed in Table 3, which, in combination with the hazard-cost pyramid in Fig. 5, provides measures to identify hazards having high mitigation potential.

6 Conclusions

This article addressed the issues of safety, resilience, and cyber security as inherent parts of the design and implementation of autonomous systems for marine surface vessels.

Development of autonomy was discussed for the strictly regulated marine area, and an architecture was suggested that could adhere to the requirements in this field.

A novel sovereign-based architecture was presented, which facilitates the requirements for safety and dependability of maritime autonomy: the ability to integrate subsystems from different vendors, testability at the system level, resilience to not-normal behaviours of instruments, dependable reactions to rarely occurring events, and resilience to possible cyber-attacks.

The article explained how behaviour-based techniques were employed to obtain diagnosis and resilience against artefacts in sensors and signals, and it addressed the assessment of risk for items and activities that cannot be accessed by direct measurement.

The concepts were exemplified by the autonomy system developed for an autonomous harbour-bus, on which the implementation will be commissioned and demonstrated in mid-2022.

Acknowledgment: The authors would like to acknowledge and express our appreciation for the great teamwork and healthy discussions among fellow colleagues within the ShippingLab Team at the Technical University of Denmark.

Funding: The authors gratefully acknowledge the sponsorship of this research by the Danish Innovation Fund through the ShippingLab project, grant 8090-00063B (the Work package on autonomy), and by the Danish Maritime Fund, Orients Fund, and the Lauritzen Fund.

References

1. Bak, T., R. Wisniewski and M. Blanke. 1996. Autonomous attitude determination and control system for the orsted satellite. In *1996 IEEE Aerospace Applications Conference. Proceedings*. IEEE.
2. Barrass, C. B. and D. R. Derrett. 2012. *Ship Stability for Masters and Mates*. Elsevier Ltd.
3. Beighton, R. – CNN. 2021. World’s first crewless, zero emissions cargo ship will set sail in Norway. https://edition.cnn.com/2021/08/25/world/yara-birkeland-norway-crewless-container-ship-spc-intl/index.html.

4. Bhamare, D., M. Zolanvari, A. Erbad, R. Jain, K. Khan and N. Meskin. 2020. Cybersecurity for industrial control systems: A survey. *Computers and Security* 89: 101677.
5. BIMCO. 2020. The Guidelines on Cyber Security Onboard Ships. Tech. rep. V4, Bimco.
6. BIMCO. 2020. First ever standard contract for autonomous ship operation underway. https://www.bimco.org/sitecore/content/bimco/home/news/contracts-and-clauses/20201106-first-ever-standard-contract-for-autonomous-ship-operation-underway.aspx.
7. Bjørn, J. V., S. Rolf and L. S. Asun. 2018. DNV GL–Maritime Remote-Controlled and Autonomous Ships. *DNV GL – Maritime*, 36 p. https://www.dnv.com/maritime/publications/remote-controlled-autonomous-ships-paper-download.html.
8. Blanke, M. 2005. Diagnosis and fault-tolerant control for ship station keeping. In: *IEEE Mediterenean Control Conf.* IEEE Xplore, 10.1109/.2005.1467217.
9. Blanke, M., M. Kinnaert, J. Lunze and M. Staroswiecki. 2016. *Diagnosis and Fault-Tolerant Control*. 3rd edition. Springer Berlin Heidelberg.
10. Blanke, M. and D. T. Nguyen. 2018. Fault-tolerant position-mooring control for offshore vessels. *Ocean Engineering* 148: 426–441.
11. Blanke, M. and M. Staroswiecki. 2006. Structural design of systems with safe behaviour under single and multiple faults. *IFAC Proceedings Volumes* 39(13): 474–479. IFAC SAFEPROCESS Symposium.
12. Blanke, M., M. Staroswiecki and N. E. Wu. 2001. Concepts and Methods in Fault-tolerant Control. In: *American Control Conference, vol. 4*. 15 p.
13. Bureau Veritas. 2019. Rules on Cyber Security for the Classification of Marine Units. Tech. rep., Bureau Veritas.
14. BV. 2019. Guidelines for Autonomous Shipping – Guidance Note NI 641 DT R01 E. Tech. rep. NI 641 DT R01 E, Bureau Veritas.
15. Chircop, A., 2018. Testing International Legal Regimes: The Advent of Automated Commercial Vessels. *The German Yearbook of International Law* 60(1):109–142.
16. Cusimano, J., M. Ayala and G. Villano. 2020. Navigating Cybersecurity Challenges in Maritime Operational Technology. https://www.maritime-executive.com/editorials/navigating-cybersecurity-challenges-in-maritime-operational-technology.
17. Dittmann, K., P. N. Hansen, D. Papageorgiou, S. Jensen, M. Lützen and M. Blanke. 2021. Autonomous surface vessel with remote human on the loop: System design for stcw compliance. *IFAC-PapersOnLine* 54(16): 224–231.
18. DNV-GL. 2018. Autonomous and remotely operated ships. Tech. rep. Class Guideline DNVGL-CG-0264, DNV-GL.
19. DnVGL. 2018. Rules for Classification – Cyber Security Notation. Tech. rep., DnVGL.
20. Earthy, J., B. S. Jones and N. Bevan. 2001. The improvement of human-centred processes – Facing the challenge and reaping the benefit of ISO 13407. *International Journal of Human Computer Studies* 55(4): 553–585.
21. Earthy, J. V. and M. Lützhöft. 2018. Autonomous ships, ICT and safety management. In: *Managing Maritime Safety*. Routledge, pp. 141–165.
22. EMSA. 2019. Annual Overview of Marine Casualties and Incidents 2014. Tech. rep., EMSA.
23. Fang, S., M. Blanke and B. J. Leira. 2015. Mooring system diagnosis and structural reliability based control for position-moored vessels. *Control Engineering Practice* 36: 12–26.
24. Fang, S., B. J. Leira and M. Blanke. 2013. Position mooring control based on a structural reliability criterion. *Structural Safety* 41: 97–106.
25. Goud, N. 2020. Cyber Attacks could easily sink Cruise Ships says Government of UK. https://www.cybersecurity-insiders.com/cyber-attacks-could-easily-sink-cruise-ships-says-government-of-uk/.
26. Goupil, P. 2011. AIRBUS state of the art and practices on FDI and FTC in flight control system. *Control Engineering Practice* 19: 524–539.
27. IEC. 2014. IEC 62065:2014 Maritime navigation and radio communication equipment and systems – Track control systems – Operational and performance requirements, methods of testing and required test results. Tech. rep., IEC.
28. IEC. 2016. IEC 61162-1:2016 Maritime navigation and radio communication equipment and systems – Digital interfaces – Part 1: Single talkers and multiple listeners. Tech. rep., IEC.
29. IMO. 2000. Resolution A.893(21) Guidelines for Voyage Planning. Tech. rep. A 2/Res.893, IMO.
30. IMO. 2007. MSC 252 (83) Adoption of the revised performance standards for INS. Tech. rep. MSC.252/Circ.83, IMO.
31. IMO. 2008. MSC.267(85) Adoption of the International Code on Intact Stability. Tech. rep. MSC.267(85), IMO.
32. IMO. 2010. The Manila Amendments to STCW. Tech. rep. STCW/CONF.2/34, IMO.
33. IMO. 2013. Guidelines for the Approval of alternatives and equivalents as provided for in various IMO instruments. Tech. rep. MSC.1/Circ.1455, IMO.
34. IMO. 2017. MSC-FAL.1-Circ.3 – Guidelines On Maritime Cyber Risk Management. Tech. rep. MSC-FAL.1-Circ.3, IMO.
35. IMO. 2018. Strategic Plan for the Organization for the Six-Year Period 2018 to 2023. Tech. rep. Resolution A.1110(30), IMO.
36. IMO. 2019. Interim Guidelines For MASS Trials. Tech. rep. MSC.1/Circ.1604, IMO.
37. IMO. 2019. Resolution A.1131(31) – List of Outputs for the 2020–2021 Biennium. Tech. rep. Resolution A.1131(31), IMO.
38. IMO. 2021. Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS). Tech. rep. MSC.1/Circ.1638, IMO.
39. ISO. 2019. ISO 11674 Ships and marine technology – Heading control systems. Tech. rep., ISO.
40. Jones, D. A. 2003. *Nomenclature for Hazard and Risk Assessment in the Process Industries, Institution of Chemical Engineers*. 2nd edition. The Institution of Chemical Engineers.
41. Leveson, N. G. 2020. *Engineering a Safer World: Systems Thinking Applied To Safety*. The MIT Press.
42. Lloyd's Register of Shipping. 1994. ATOMOS II – Hazard Identification.
43. Manley, J. E. 2019. Waypoints on the voyage to autonomous ships. In: *OCEANS 2019 MTS/IEEE SEATTLE*. IEEE.
44. MUNIN. 2016. Maritime Unmanned Navigation through intelligence in networks. Tech. rep., DnV.
45. Nissov, M. C., D. Dagdilelis, R. Galeazzi and M. Blanke. 2021. Analyzing cyber-resiliency of a marine navigation system from behavioral relations. In: *Proc. European Control Conference ECC'2021*. IEEE Xplore.
46. NIST. 2021. Developing Cyber-Resilient Systems:A Systems Security Engineering Approach. Tech. rep. 800-160, National

Institute of Standards and Technology.
47. Porathe, T., J. Prison and Y. Man. 2014. Situation awareness in remote control centres for unmanned ships. In *Human Factors in Ship Design and Operation, vol. 27*. pp. 1–9.
48. Rokseth, B., O. I. Haugen and I. B. Utne. 2019. Safety Verification for Autonomous Ships. *MATEC Web of Conferences* 273: 02002.
49. Rolls-Royce. 2016. Remote and Autonomous Ship – The next steps. Tech. rep., Rolls-Royce.
50. Shiokari, M. and S. Ota. 2019. Considerations on the regulatory issues for realization of Maritime Autonomous Surface Ships. *Journal of Physics: Conference Series* 1357(1): 012005.
51. Thieme, C. A. 2018. Risk Analysis and Modelling of Autonomous Marine Systems. PhD thesis, Norwegian University of Science and Technology, Trondheim.
52. Thieme, C. A., A. Mosleh, I. B. Utne and J. Hegde. 2020. Incorporating software failure in risk analysis – Part 1: Software functional failure mode classification. *Reliability Engineering and System Safety* 197: 106803.
53. UN. 1982. United Nations Convention on the Law of the Sea. Tech. rep., United Nations.
54. Vander Maelen, S., M. Buker, B. Kramer, E. Bode, S. Gerwinn, G. Hake and A. Hahn. 2019. An Approach for Safety Assessment of Highly Automated Systems Applied to a Maritime Traffic Alert and Collision Avoidance System. In: *2019 4th International Conference on System Reliability and Safety, ICSRS 2019*. IEEE, pp. 494–503.
55. Veal, R. and H. Ringbom. 2017. Unmanned ships and the international regulatory framework. *Journal of International Maritime Law* 23(2): 100–118.

Bionotes

Kjeld Dittmann
Automation and Control Group, Department of Electrical and Photonics Engineering,
Technical University of Denmark,
Kgs. Lyngby, Denmark
kjeditt@elektro.dtu.dk

Kjeld Dittmann is a Post Doc. with the Department of Electrical and Photonics Engineering at the Technical University of Denmark (DTU). With research interests in networked control systems and cyber-physical systems with applications in transportation, energy, and automation networks, Kjeld Dittmann serves as Chair of the ShippingLab programme.

Mogens Blanke
Automation and Control Group, Department of Electrical and Photonics Engineering,
Technical University of Denmark,
Kgs. Lyngby, Denmark
mb@elektro.dtu.dk

Mogens Blanke is Professor in Automation and Control at the Technical University of Denmark and Principal Investigator for the ShippingLab research effort on autonomy. The main fields of research are diagnosis, prognosis, and fault tolerant autonomous systems.