Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P1792-cd

Towards real-time safety monitoring for autonomous inland waterway vessels: The SeaGuard tool

Konstantinos Louzis

School of Naval Architecture & Marine Engineering, National Technical University of Athens, Greece. E-mail: klouzis@mail.ntua.gr

Alexandros Koimtzoglou

School of Naval Architecture & Marine Engineering, National Technical University of Athens, Greece. E-mail: akoim@mail.ntua.gr

Marios Koimtzoglou

School of Naval Architecture & Marine Engineering, National Technical University of Athens, Greece. E-mail: marioskoim@mail.ntua.gr

Panagiotis Katsos

School of Naval Architecture & Marine Engineering, National Technical University of Athens, Greece. E-mail: pkatsos@mail.ntua.gr

Nikolaos P. Ventikos

School of Naval Architecture & Marine Engineering, National Technical University of Athens, Greece. E-mail: niven@deslab.ntua.gr

Autonomous operation has the potential to significantly enhance inland waterway transport by facilitating a shift to zero-emission propulsion and contributing to the competitiveness to alternative transport modes like road and rail. Autonomous vessels integrate hardware, advanced digital and software systems, as well as humans-in-the-loop and therefore constitute complex Socio-Technical Systems, whose safety can be affected by random faults, as well as vulnerabilities to intentional cyber-attacks. Despite technological advancements that allow for crewless or remotely controlled vessels, autonomous or remote control needs to be enhanced with risk awareness to ensure that associated uncertainties can be managed in real-time, and that autonomous operation is both safe and resilient. To address these challenges, the EC-funded, Horizon Europe project AUTOFLEX (AUTOnomous and small FLEXible vessels) develops the SeaGuard tool, which is intended to perform real-time monitoring and risk assessment given faults, unsafe system interactions, and cyber-security threats, with the aim to facilitate reverting to a safe state within a specified time window by proposing appropriate risk control measures in the form of decision support to operators and relevant stakeholders. To achieve this, SeaGuard integrates detection of anomalies either in the form of cyber-attacks or faults with real-time risk assessment and evaluation of candidate risk control measures. This paper describes the functions required for SeaGuard to accomplish its objectives, the approach that will be implemented for assessing the safety level, as well as a high-level overview of the supporting methodological framework. SeaGuard is expected to significantly contribute to the feasibility of autonomous operations in inland waterways and by extension to the competitiveness of this transport mode against land-based transportation.

Keywords: Autonomous vessels, Inland Waterways Transport, Safety assessment, Real-time monitoring, Resilience.

1. Introduction

In alignment with the strategy of the European Union (EU) to address the challenges related to reducing the environmental impact of transportation, as well as the impact of externalities, such as congestion, by shifting cargoes from land-based to water-based transportation modes (Essen et al. 2019), there is increased interest in enhancing the use of Inland

Waterways (IWW). However, IWW vessels have currently higher external costs per ton of transported cargo compared to land-based modes (Essen et al. 2019). In this context, there is a need to modernize the IWW fleet towards zero-emission propulsion, which can be facilitated by autonomous operation as removing human-centred design features (i.e. the bridge) may free space for batteries in the case of fully electric propulsion without compromising cargo carrying capacity (Bačkalov et al. 2024). However, despite the technological advancements that enable crewless or remotely controlled operation, there are uncertainties with respect to new and emerging risks autonomous ships may be exposed to, such as cyber-attacks that may lead to unavailability or malfunctioning of critical systems, such as propulsion (Bolbot et al. 2020).

To address these challenges, autonomous ships require sophisticated frameworks for continuous monitoring and dynamic risk evaluation especially with increasing levels of autonomy and less human involvement (Johansen et al. 2023). Such frameworks can support the development of risk-based supervisory controllers that consider both safety and cyber-security (Utne et al. 2020). Although there have been developments in the field of real-time risk assessment for autonomous ships, there are challenges related to updating the supporting hazard analyses to consider new and emerging risks, the dependence on expert knowledge due to lack of historical data, as well as the limited integration of safety and cyber-security.

The research problem addressed by our work is how appropriate risk control measures can be determined in real-time based on an estimate of the risk resulting from anomalies during the operation of autonomous ships. The objective of this paper is to describe the functionalities and methodological framework of the SeaGuard tool that will be developed for supporting the operation of a conceptual IWW vessel designed within the context of the EC-funded research project AUTOFLEX. This vessel will be, at least periodically, unmanned, will carry containers, and will be equipped with battery-powered electric propulsion and azimuth thrusters (Bačkalov et al. 2024). SeaGuard addresses the identified challenges through a real-time risk assessment framework that integrates fault detection with intrusion detection, which evaluate deviations from the norm without relying on historical data, with system and risk models that are jointly used for assessing the effect of reduced system capabilities to the likelihood of unwanted consequences. The relationship between cyber-attacks and faults is considered in determining the impact on the system's functionality.

The rest of this paper is structured as follows: Section 2 reviews the literature related to methodologies used for real-time risk and cyber-risk assessment. Section 3 describes the functionalities and underlying methodological framework of SeaGuard. Section 4 briefly discusses the novelty of our contribution. The paper concludes with a description of the next steps of our research.

2. Related Work

Models that aim to assess risks in real-time should use various sources of data, including sensor output, reflect dynamic changes in operation, including the impact of changes in subsystems or components, be updated when new information is available, identify Risk Influencing Factors (RIFs) that need to be monitored and model their impact on the risk level, as well as take uncertainty into account (Yang and Utne 2022).

Our review of the literature on maritime real-time risk models shows that static and dynamic Bayesian Networks (BNs) are mainly used differing in terms of how the included risk factors are identified either exploiting the results of formal hazard analyses or based on information from literature reviews and expert knowledge. Spahic et al. (2023) have also highlighted the importance of providing contextual information resulting from risk analyses in real-time anomaly detection applications.

Utne et al. (2020) developed a framework that uses the context provided by the Systems Theoretic Process Analysis (STPA) to create risk models in the form of static BNs that can be updated with real-time data depending on the frequency this data is available. The model is structured to reflect the following sequence: Input RIF, High-level RIF, UCA, and System-level hazard. High-level RIFs are identified from the scenarios resulting from STPA and Input RIFs are indicators that can be measured in real-time and are related to causal factors included in the scenarios. This framework was extended by

Johansen and Utne (2022) to include consequences as a result of system-level hazards and worst-case environmental conditions in the BN. The output is exploited in the form of a "risk cost" in a cost function that is subsequently used by a controller to make decisions during operation. Following a similar approach, Basnet et al. (2023) implemented techniques, such as Noisy-OR/MAX gates, Parent-Divorcing, and modular BN, to reduce the number of BN nodes that resulted from the STPA.

Luo et al. (2024) developed a Dynamic Bayesian Network (DBN) that integrates static and dynamic risk factors, i.e. factors that are time-dependent, that affect the real-time risk level during the berthing operation of MASS. The contextual information for the development of the risk model resulted from the application of Task Failure Modes and Effects Analysis (Task FMEA), as well as by the relevant literature and expert knowledge. In a less structured approach, Kristensen et al. (2022) developed a DBN suitable for Dynamic Risk Assessment (DRA) based on information derived from literature reviews that focuses on the effect of inadequate Situation Awareness (SA) and loss of power on the mission performance of an Unmanned Surface Vessel (USV).

A different approach has been followed by Zeleskidis et al. (2020) in the domain of railway safety who use the results of STPA to develop a mathematical model in the form of an acyclic diagram that aims to assess the safety level of a system in real-time, which is characterized by the time remaining until accident occurrence and how close the system is to experiencing a loss.

The models reviewed in the maritime risk domain have mainly been demonstrated in case studies that involve ship-related technical (e.g. system state and reliability), as well as environmental (e.g. weather, sea state, maritime traffic) risk factors that affect navigational risks (i.e. collision and grounding). Gomola and Utne (2024) focused on software controller failures by extending the STPA framework with the integration of the Systems Modelling Language (SysML) and used a case study involving the navigation and collision avoidance system of a semi-autonomous ferry. Although other types of risk factors, such as human-related and cyber-security issues, have not been included in the case studies, the methodologies themselves have no limitation in this regard. In fact, Yang and Utne (2022) have highlighted the need for an online risk model to consider (cyber)security issues.

Frameworks for offline cyber-risk assessment tailored specifically to autonomous ships have been developed, such as Maritime Cyber-Risk Assessment (MaCRA) that is a model-based approach that considers attacker profiles, ease of exploitation, and potential rewards (Tam and Jones 2019). Techniques from other domains have also been applied to autonomous ship architectures, such as Microsoft's STRIDE threat modelling methodology (Kavallieratos, Katsikas, and Gkioulos 2019). However, these approaches do not address the relationship between safety and cyber-security.

Bolbot et al. developed a structured approach for hazard identification that integrates risk assessment with cyber-security and security assessments in the initial design phase for autonomous IWW vessels (2021). The assessments are not done in parallel but are integrated into a single process. However, the authors propose conducting a dedicated cyber-risk assessment to reduce the uncertainty related to the experts' risk ranking, such as the CYber-Risk Assessment for Marine Systems (CYRA-MS) developed by Bolbot et al. (2020), which is based on the Cyber Preliminary Hazard Analysis (CPHA) and also considers safety-related consequences of cyber-attacks.

Amro et al. investigated how cyber-attacks may trigger failures through the Six Step Model (SSM), which is based on the GTST-MLD notation and facilitates examining the interrelationship between safety and cyber-security issues (2020). Cyber-attacks are identified from the implementation of the STRIDE methodology and failures from a Preliminary Hazard Analysis (PHA). The authors have not considered the consequences but identified safety and cyber-security countermeasures and the relationship between them based on how failures and attacks impact the functionalities of the system.

Such frameworks can support the development of real-time cyber-risk assessments, similarly to how hazard analyses support real-time risk assessments, combined with an Intrusion Detection System (IDS). Detecting cyber-attacks in real-time can accomplished by different

Machine Learning (ML) techniques, such as Deep Neural Networks (DNN) (Thirimanne et al. 2022), as well as classifiers such as Naïve Bayes (NB), Random Forest (RF), and Decision Tree (DT) that are trained and evaluated using benchmark datasets, e.g. KDD Cup 99 (Alqahtani et al. 2020).

A promising class of algorithms for application in anomaly detection (see Bayar et al. 2015), as well as intrusion detection are Artificial Immune Systems (AIS) that are inspired by different theories about how the biological immune system identifies and responds to harmful microorganisms. Indicatively, Shen and Wang (2011) proposed an IDS based on the Negative Selection Algorithm (NSA). Pinto et al. (2022) developed a method for detecting anomalies in Cyber-Physical Systems (CPS) using a novel variant of the Dendritic Cell Algorithm (DCA). Wang et al. (2016) proposed a multi-class classifier based on negative selection, to distinguish between the following four operational modes of mining equipment: normal, transition, abnormal, and danger.

3. The SeaGuard framework

3.1. *Objectives and functions*

The functionalities of SeaGuard comprise the following three main groups that are accomplished sequentially (Fig. 1): 1) detection and identification, 2) risk estimation, and 3) risk control and mitigation.

The first objective of SeaGuard will be to detect anomalies with reference to the system's "normal" or expected behaviour, as described by its operational envelope (see Fjørtoft and Holte 2022). The detection will rely on gathering information in real-time from various system components (e.g. sensors that contribute to situation awareness, network traffic) and using it to quantify indicators of "symptoms" of irregular system behaviour. Such anomalies may either be attributed to random faults (e.g. sensor drift or saturation, or a fault in the thruster), or cyber-attacks (e.g. high volume of network traffic from an external source that overloads the onboard network). SeaGuard will determine the likelihood of each possibility and subsequently identify which part of the system has been affected.

The second objective of SeaGuard will be to prioritise the detected anomalies based on how they can propagate through the system and compromise its capability to achieve its intended functions. In this context, cascading effects between cyber-attacks and faults will also be assessed. For example, given high confidence that a cyber-attack is occurring, SeaGuard will assess the likelihood of resulting faults (e.g. in case critical navigation systems become unresponsive due to overload of the onboard network). Given high confidence that the anomaly is due to a random fault, SeaGuard will assess the likelihood of the system being more vulnerable to cyber-attacks.

Prioritization of anomalies will be accomplished using a risk metric that will take into account the reduction of the system's capabilities and the severity of the consequences given the detected anomaly, as well as the associated uncertainties and background knowledge in the general form of a (C', Q, K) triplet (Aven et al. 2018). SeaGuard will consider both cyber-security (e.g. data theft) and safety (e.g. grounding or collision accidents) related consequences either as a result of a cyber-attack or a fault.

The third objective of SeaGuard will be to identify applicable mitigation measures depending on whether the anomaly is attributed to a fault or a cyber-attack and with the aim to either restore the system's functionalities (partly or fully) or minimise risk given that normal operation cannot be resumed. These measures will be prioritized based on their expected effectiveness that will be defined with specific criteria. For example, in case the onboard network traffic is unusually high due to an external actor, rate limiting at the ship's firewall would prevent further congestion in the network, and the network routers could be configured to block requests to broadcast addresses.

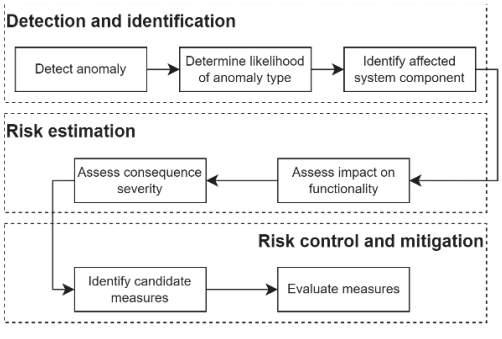

Fig. 1. The functionalities of SeaGuard.

3.2. *Methodology*

The methodological framework for SeaGuard will integrate data-driven Machine Learning (ML) models designed for real-time detection of operational anomalies, probabilistic models for real-time risk assessment, and deterministic models for real-time response (Fig. 2). This approach ensures that the system will be adaptive and responsive to evolving threats within the operational environment of inland autonomous vessels.

SeaGuard will simultaneously detect and identify cyber-attacks and random faults that manifest as anomalies in network traffic data and in the data produced by different system components that are essential for the autonomous operation of the vessel (e.g. RADAR, GPS etc.) respectively. This parallel approach provides robustness and comprehensiveness to safeguarding critical onboard systems. Detecting fault-related anomalies will be accomplished by using data available in real-time to quantify derivative variables that describe the system state and from which "symptoms" of abnormal operation may be identified. The system state estimator will feed this information into an algorithm based on the principles of Artificial Immune Systems (AIS), such as the Negative Selection Algorithm (NSA). This data AIS (dAIS) will compare how the system currently functions against the normal operational envelope, as well as identify which system component is the source of the fault.

Detecting anomalies associated with cyber-attacks will be accomplished by a two-layer approach that will monitor network-wide and component-specific traffic. The first will involve a cyber-attack classifier to identify relevant patterns in network traffic between the vessel's Local Area Network (LAN) and its network gateway, creating a critical monitoring layer that isolates sensitive internal communications from external threats. Similarly to the work by Thirimanne et al. (2022), a real-time feature extraction mechanism will extract data from the network traffic and provide input to a multi-class classifier, such as a Deep Neural Network (DNN), which will detect and classify the cyber-attack with specific probability scores. Training of the model will be done using benchmark datasets for intrusion detection, such as the KDD Cup 99 and its refined version NSL-KDD, which simulate network traffic with both normal and malicious activities.

For monitoring component-specific cyber-attacks, each component will be connected to dedicated network traffic feature extractors that will be integrated into the gateway. The purpose of these extractors will be to filter the data packets generated by each component, extract relevant features, and provide the output to a dedicated traffic AIS (tAIS). The tAIS will perform anomaly detection ensuring that deviations from normal behaviour are flagged in real time and that specific components being attacked can be identified. The output from the multi-class classifier will be concatenated with the output of the tAIS to increase confidence in the detection of a cyber-attack. By leveraging both tAIS and dAIS, the aim is to identify the source of the detected anomaly as either a cyber-attack or a fault and consequently achieve comprehensive severity estimation and effect assessment on the ship's critical components.

After identifying the affected component, the SeaGuard Risk Estimator will first determine how the system's capabilities will be reduced considering cascading effects by using a model that describes both functional and structural aspects of the system, such as the one used by Amro et al. (2020) or a model similar to the control structure used in STPA. This will be used as input to a risk model that will describe the relationship between system capabilities and consequences in probabilistic terms, such as a Bayesian Network.

The SeaGuard Risk Control Engine will propose appropriate mitigation measures based on a pre-defined library of candidates, which will be tailored to the vessel's operational requirements, depending on the type of the detected anomaly and the criticality as indicated by the result of the risk model. To ensure rapid and effective response, SeaGuard will use specific criteria to select the most appropriate candidate, such as minimizing disruptions to the vessel's autonomous functions and risk reduction effectiveness.

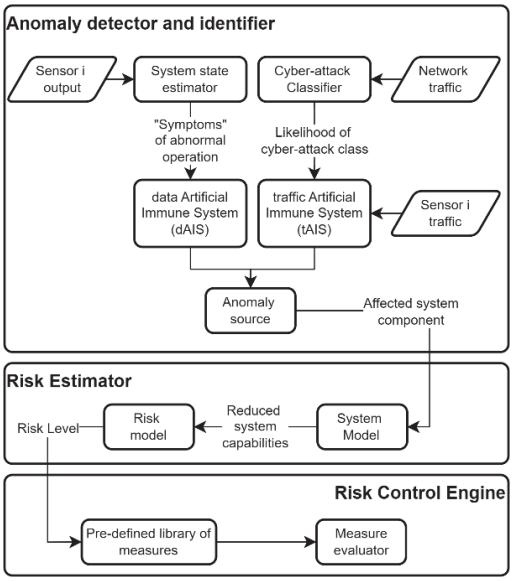

Fig. 2. The methodological framework of SeaGuard.

4. Discussion

SeaGuard will be designed to accomplish the following functionalities: 1) detection and identification, 2) risk estimation, and 3) risk control. The underlying methodological framework presented in this paper is expected to enhance vessel resilience in terms of anticipating disturbances, monitoring performance, and responding to hazards, which are the elements of resilient systems according to Hollnagel et al. (2006).

SeaGuard is expected to be suitable for providing real-time risk awareness as it fulfils the requirements for online risk models identified by Yang and Utne (2022). It will use various data sources to obtain information in real-time to assess the impacts of detected anomalies on the functionality of the system. Furthermore, SeaGuard will consider the effect of RIFs in the detection and identification stage, which will be reflected in the variables used for estimating the system state, as well as in the risk estimation stage, which will be reflected in the structure of the risk model. The methodological approach of SeaGuard includes uncertainties that relate to the type and source of the detected anomaly given the real-time information that will be used, as well as the background knowledge supporting the models used to assess the reduction in system capabilities and the severity of the consequences. To increase confidence in the effectiveness of the proposed risk control measures, our methodological framework needs to include appropriate characterization of uncertainties, in alignment with the discussion in Aven et al. (2014), adapted for the real-time assessment setting. Although a detailed discussion on this aspect is outside the scope of this paper, a hybrid approach may be preferred, which will involve complementing the probabilistic descriptions with other measures of uncertainty. Indicatively, a possible approach would be to combine Dempster-Schafer's Theory of evidence for the anomaly detection with a semi-quantitative approach for evaluating the strength of background knowledge supporting the models in the SeaGuard Risk Estimator (see Aven 2013).

The reviewed methods for real-time risk assessment use information as evidence in BNs that model the causal relationships between RIFs to update the probability of accident occurrence and their consequences. As the risk models are created from contextual information provided by different hazard analysis techniques, risk is assessed by updating the probabilities of pre-defined hazardous scenarios in real-time. Furthermore, the probabilistic relationships are quantified with available historical data and expert knowledge where data is not available.

Updating the probabilities of pre-defined scenarios makes including new and emerging risks challenging considering that the hazard analysis needs to be updated offline (see Escande, Proust, and Le Coze 2016). By integrating real-time anomaly detection with risk assessment, SeaGuard aims to address this by effectively creating scenarios given a triggering event in real-time through by combining a system model with a risk model. However, SeaGuard will rely to some extent on offline hazard analyses to provide context for the anomaly detection, which aims to address the shortcomings of purely data-driven approaches, i.e. providing too many false positives (noise) and false negatives (overlooking safety critical anomalies), as shown by Spahic et al. (2023).

In offline risk assessment frameworks that consider both cyber-security and safety, the relationship between the two is treated in terms of: 1) how cyber-attacks can lead to safety-related consequences (see Bolbot et al. 2021), and 2) how cyber-attacks may trigger failures (see Amro et al.

2020). SeaGuard will consider these types of relationships and will also consider how faults affect the likelihood of making the system vulnerable to cyber-attacks by integrating intrusion detection with real-time risk assessment.

The immune paradigm was selected as the basis for SeaGuard's anomaly detection due to the benefits of this class of algorithms in terms of the data required for training and their classification performance compared to other types of classifiers. AIS algorithms typically only require data that describe normal operation for training (see Bayar et al. 2015), which makes them suitable for cases where historical data are not widely available, such as the case of autonomous ships. Furthermore, using this approach in real-time implies that a set of circumstances is evaluated as "hazardous" not solely through the probability of leading to an accident, but also from the dissimilarity to how the system is normally expected to operate. In terms of performance, the literature indicates that AIS algorithms perform with higher classification accuracy compared to Neural Networks (NN) and Support Vector Machines (SVM) (Wang et al. 2016), as well as compared to other anomaly detection techniques, such as deep neural networks and regression approaches, although in some cases false alarm rates may be higher (Pinto, Pinto, and Gonçalves 2022).

5. Conclusion

The SeaGuard tool aims to provide real-time risk awareness to the autonomous operation of IWW vessels by detecting anomalies either due to cyber-attacks or random faults, identifying which system component has been affected, assessing how the anomaly can impact the system's functionality and subsequently the likelihood of unwanted consequences. Based on the estimated risk and the anomaly type, SeaGuard will provide decision support in the form of proposing the most suitable risk control measure.

Methodologically, SeaGuard addresses the challenges related to real-time risk assessment by integrating anomaly detection with risk assessment without relying on a set of pre-defined hazardous scenarios, exploiting immune-inspired classifiers that evaluate similarity to a norm without relying on historical data, and by considering the cascading effects on the system's functionality between cyber-attacks and faults.

Future research steps include complementing the framework with techniques to characterize uncertainty, developing and testing the algorithms and models that will be integrated through a suitable simulation framework. Finally, SeaGuard is expected to significantly contribute to ensuring safe and resilient autonomous operation of IWW vessels and consequently unlocking their potential towards minimizing the external costs of transportation.

Acknowledgments

The paper presents the outcomes of research conducted within the framework of the project AUTOFLEX (AUTOnomous small and FLEXible vessels) which received funding from the European Union Horizon Europe Programme under the Grant Agreement 101136257.

References

Alqahtani, H., I. H. Sarker, A. Kalim, S. M. M. Hossain, S. Ikhlaq, and S. Hossain. 2020. 'Cyber Intrusion Detection Using Machine Learning Classification Techniques'. In *Computing Science, Communication and Security*, edited by Nirbhay Chaubey, Satyen Parikh, and Kiran Amin, 121–31. Singapore: Springer.

Amro, A., G. Kavallieratos, K. Louzis, and C. A. Thieme. 2020. 'Impact of Cyber Risk on the Safety of the MilliAmpere2 Autonomous Passenger Ship'. *IOP Conference Series: Materials Science and Engineering* 929 (November):012018.

Aven, Terje. 2013. 'Practical Implications of the New Risk Perspectives'. Reliability Engineering and System Safety 115:136–45.

Aven, T., P. Baraldi, R. Flage, and E. Zio. 2014. *Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods*. 1 edition. Chichester, West Sussex, United Kingdom: Wiley.

Aven, T., Y. Ben-Haim, H. Boje Andersen, T. Cox, E. López Droguett, M. Greenberg, Seth Guikema, et al. 2018. 'Society for Risk Analysis Glossary'.

Bačkalov, I., H-C. Burmeister, L. Isidorović, J. Jasa, M. Josipović, K. Kloch, A. Koimtzoglou, et al. 2024. 'Impact of Automation and Zero-Emission Propulsion on Design of Small Inland Cargo Vessels'. *Journal of Physics: Conference Series* 2867 (1): 012018.

Basnet, S., A. BahooToroody, M. Chaal, J. Lahtinen, V. Bolbot, and O. A. Valdez Banda. 2023. 'Risk Analysis Methodology Using STPA-Based Bayesian Network- Applied to Remote Pilotage

Operation'. *Ocean Engineering* 270 (February):113569.

Bayar, N., S. Darmoul, S. Hajri-Gabouj, and H. Pierreval. 2015. 'Fault Detection, Diagnosis and Recovery Using Artificial Immune Systems: A Review'. *Engineering Applications of Artificial Intelligence* 46:43–57.

Bolbot, V., G. Theotokatos, L. A. Wennersberg, J. Faivre, D. Vassalos, E. Boulougouris, Ø. J. Rødseth, P. Andersen, A-S. Pauwelyn, and A. Van Coillie. 2021. 'A Novel Risk Assessment Process: Application to an Autonomous Inland Waterways Ship'. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 1–23.

Bolbot, V., G. Theotokatos, E. Boulougouris, and D. Vassalos. 2020. 'A Novel Cyber-Risk Assessment Method for Ship Systems'. *Safety Science* 131 (November):104908.

Escande, J., C. Proust, and J. C. Le Coze. 2016. 'Limitations of Current Risk Assessment Methods to Foresee Emerging Risks: Towards a New Methodology?' *Journal of Loss Prevention in the Process Industries* 43 (September):730–35.

Essen, H., L. Wijngaarden, A. Schroten, D. Sutter, C. Bieler, S. Maffii, M. Brambilla, D. Fiorello, F. Fermi, and R. Parolin. 2019. 'Handbook on the External Costs of Transport, Version 2019'. 18.4 K83. 131.

Fjørtoft, K., and E. Holte. 2022. 'Implementing Operational Envelopes for Improved Resilience of Autonomous Maritime Transport'. In *Human Factors in Transportation. AHFE (2022) International Conference*, edited by Katie Plant and Gesa Praetorius. Vol. 60. USA: AHFE Open Access.

Gomola, A., and I. B. Utne. 2024. 'A Novel STPA Approach to Software Safety and Security in Autonomous Maritime Systems'. *Heliyon* 10 (10).

Hollnagel, E., D. D. Woods, and N. G. Leveson. 2006. *Resilience Engineering: Concepts and Precepts*. Burlington, USA: Ashgate Publishing Company.

Johansen, T., S. Blindheim, T. R. Torben, I. B. Utne, T. A. Johansen, and A. J. Sørensen. 2023. 'Development and Testing of a Risk-Based Control System for Autonomous Ships'. *Reliability Engineering & System Safety*, February, 109195.

Johansen, T., and I. B. Utne. 2022. 'Supervisory Risk Control of Autonomous Surface Ships'. *Ocean Engineering* 251 (May):111045.

Kavallieratos, G., S. Katsikas, and V. Gkioulos. 2019. 'Cyber-Attacks Against the Autonomous Ship'. In *Computer Security*, edited by S. Katsikas, F. Cuppens, N. Cuppens, C. Lambrinoudakis, A. Antón, S. Gritzalis, J. Mylopoulos, and C. Kalloniatis, 20–36. Cham: Springer International Publishing.

Kristensen, S. D., Y. Liu, and I. B. Utne. 2022. 'Dynamic Risk Analysis of Maritime Autonomous Surface Ships'. In *Probabilistic Safety Assessment and Management PSAM 16, June 26-July 1*. Honolulu, Hawaii, USA.

Luo, X., H. Ling, M. Xing, and X. Bai. 2024. 'A Dynamic-Static Combination Risk Analysis Framework for Berthing/Unberthing Operations of Maritime Autonomous Surface Ships Considering Temporal Correlation'. *Reliability Engineering & System Safety* 245 (May):110015.

Pinto, C., R. Pinto, and G. Gonçalves. 2022. 'Towards Bio-Inspired Anomaly Detection Using the Cursory Dendritic Cell Algorithm'. *Algorithms* 15 (1): 1.

Shen, J., and J. Wang. 2011. 'Network Intrusion Detection by Artificial Immune System'. In *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, 4716–20.

Spahic, R., V. Hepsø, and M. A. Lundteigen. 2023. 'A Novel Warning Identification Framework for Risk-Informed Anomaly Detection'. *Journal of Intelligent & Robotic Systems* 108 (2): 17.

Tam, K., and K. Jones. 2019. 'MaCRA: A Model-Based Framework for Maritime Cyber-Risk Assessment'. *WMU Journal of Maritime Affairs* 18 (1): 129–63.

Thirimanne, S. P. L. Jayawardana, L. Yasakethu, P. Liyanaarachchi, and C. Hewage. 2022. 'Deep Neural Network Based Real-Time Intrusion Detection System'. *SN Computer Science* 3 (2): 145.

Utne, I. B., B. Rokseth, A. J. Sørensen, and J. E. Vinnem. 2020. 'Towards Supervisory Risk Control of Autonomous Ships'. *Reliability Engineering & System Safety* 196 (April):106757.

Wang, Z., X. Xu, L. Si, R. Ji, X. Liu, and C. Tan. 2016. 'A Dynamic Health Assessment Approach for Shearer Based on Artificial Immune Algorithm'. *Computational Intelligence and Neuroscience* 2:12.

Yang, R., and I. B. Utne. 2022. 'Towards an Online Risk Model for Autonomous Marine Systems (AMS)'. *Ocean Engineering* 251 (May):111100.

Zeleskidis, A., I. M. Dokas, and B. Papadopoulos. 2020. 'A Novel Real-Time Safety Level Calculation Approach Based on STPA'. In *MATEC Web of Conferences*, 314:01001. EDP Sciences.