Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Risk Analysis on Autonomous Vessels based on Systems Theory – Application of NET-HARMS method

To cite this article: Alexandros Koimtzoglou *et al* 2024 *J. Phys.: Conf. Ser.* **2867** 012046

Risk Analysis on Autonomous Vessels based on Systems Theory – Application of NET-HARMS method

Alexandros Koimtzoglou[1*], Nikolaos P. Ventikos[1], Dimitrios Routsis[1], Konstantinos Louzis[1]

[1] School of Naval Architecture & Marine Engineering, National Technical University of Athens, Greece.
Heroon Polytechniou 9, Zographou Campus, Athens 15780, Greece
*Corresponding author: akoim@mail.ntua.gr

Abstract. The shift towards Maritime Autonomous Surface Ships is a significant development in the maritime logistics industry, with the potential to enhance efficiency, safety, and environmental sustainability. However, the integration of autonomous systems also presents new challenges and risks, particularly in the absence of empirical data for traditional risk assessment methodologies. This research tackles this problem by utilizing the Net-worked Hazard Analysis and Risk Management System (Net-HARMS) method, a systems thinking method that hasn't been previously employed in examining MASS. The method analyses the risks associated with the EC-funded, H2020, MOSES Project, which included a concept for automating the manoeuvring and docking processes with autonomous tugboats. The Net-HARMS method offers a comprehensive and holistic approach to risk assessment, overcoming the limitations of conventional probabilistic models. By constructing a Hierarchical Task Analysis and a task network, the research maps the system's operational framework and explores task interdependencies. The use of a risk mode taxonomy allows for the identification of task-specific and emergent risks, which are then assessed by utilising the risk matrix of the Risk-Based Assessment Tool developed by DNV, to assess the final risk as a function of the effectiveness of each risk mitigation layer and the severity of the identified task consequences. The findings provide valuable insights into critical tasks requiring enhanced risk control measures and contribute to the development of safety constraints necessary for the successful implementation of autonomous shipping technologies. By applying Net-HARMS method to the realm of autonomous ships, this research not only fills a significant gap in maritime risk analysis but also sets a precedent for future studies in this rapidly evolving field.

1. Introduction

The concept of autonomous transportation systems emerged several decades ago but has only recently become a reality, driven by technological advancements [1], [2]. In the maritime sector, the digital revolution of the 1970s marked the beginning of computerized ship control, evolving into today's "Shipping 4.0" or "Maritime 4.0" era, characterized by cyber-physical systems and increased autonomy [3], [4]. The push towards autonomy in maritime transport is driven by multiple factors, including enhanced safety standards, reduced freight costs, environmental sustainability, and addressing the seafarer shortage [2]. Autonomous ships promise operational efficiency, reduced human error, and increased cargo capacity by eliminating crew

Journal of Physics: Conference Series **2867** (2024) 012046 doi:10.1088/1742-6596/2867/1/012046

accommodations. Despite these advantages, fully autonomous vessels are still in the experimental phase, although various levels of automation are already employed, especially in container shipping [5].

Risk analysis for Maritime Autonomous Surface Ships (MASS) involves identifying, assessing, and mitigating potential hazards associated with autonomous operations. The existing methods for risk assessment of MASS face significant challenges due to the lack of practical application and limited data availability. Traditional risk assessment models rely heavily on historical data, and expert judgement to build their input for modelling and quantification. However, the novelty of autonomous ships means that relevant historical data is sparse, and the hazards leading to disruptive events are highly uncertain and difficult to model quantitatively [6]. Moreover, the complexity of autonomous ship systems, which incorporate new hardware, Artificial Intelligence (AI) systems, and human-operator interactions, introduces hazards distinct from those in conventional ships. The consequences of similar accidents can also differ between conventional and autonomous ships. Consequently, many attempts to apply conventional ship data to assess the risks of autonomous ships result in invalid risk assessments. This inadequacy highlights the necessity of adopting systems thinking methods, which are more suitable for fields in their early stages of development. Systems thinking can address these complexities by providing a comprehensive approach to understanding the interactions between system components, including human roles [7].

Therefore, the research presented in this paper deals with this aspect, as it suggests the application of the novel system thinking method Net-worked Hazard Analysis and Risk Management System (NET-HARMS) on a MASS system, the autonomous tugboat concept of MOSES. Furthermore, it combines NET-HARMS outcomes with the risk matrix proposed in the Risk-Based Assessment Tool (RBAT) Study [8] in order to assess the risk of the emergent risk consequences identified from the application of NET-HARMS. The remaining sections of the paper are structured as follows: The main projects of MASS, the proposed autonomy levels of MASS, and a literature review regarding risk analysis in MASS are briefly summarised in Section 2. The research methodology is analysed in Section 3. Then in Section 4, the application of the suggested methodology on the MOSES project is highlighted along with the results. Finally, to outlining future research directions, the paper concludes with insights on the application of NET-HARMS in the risk analysis for MASS.

2. Literature review

Key projects like MUNIN, ReVolt, AAWA, and YARA Birkeland illustrate the global efforts to develop autonomous vessels, highlighting the integration of advanced technologies and the potential for fully autonomous operations. These initiatives reflect a broader trend towards automation in maritime transport, driven by both economic and environmental considerations [9], [10].

The International Maritime Organization (IMO) and classification societies like DNV, Lloyd's Register, and Bureau Veritas have proposed various autonomy levels for MASS. These range from ships with automated processes and decision support to fully autonomous ships capable of independent decision-making. The classification schemes reflect the complexity of integrating autonomy into maritime operations, balancing human involvement with technological capabilities [11], [12], [13], [14].

Risk analysis and hazard identification are inseparable parts of MASS risk management. Systems thinking is a crucial tool in hazard identification and risk analysis on MASS because it

Journal of Physics: Conference Series **2867** (2024) 012046 doi:10.1088/1742-6596/2867/1/012046

provides a holistic approach to understanding the increased complexity of the system. The System-Theoretic Process Analysis (STPA), Event Analysis of Systemic Teamwork - Broken Links (EAST-BL), and Net-HARMS are the three most widely used system-thinking methods for hazard identification [15]. However, in the maritime field, STPA has been mainly applied to MASS systems for hazard identification, EAST-BL has been tested only once [15] and the first attempt to utilize NET-HARMS in the risk identification process for a MASS system is presented in this study.

Compared to other systems thinking methods, NET-HARMS extends beyond traditional sharp-end focused risk assessment techniques by identifying risks across the entire system[7]. A significant advantage of NET-HARMS is its ability to identify both task risks and emergent risks. Based on Systematic Human Error Reduction and Prediction Approach (SHERPA), which boasts strong reliability and validity evidence [16], NET-HARMS is easy to learn, apply, and requires minimal training. Its generic nature makes it applicable in any domain, and creating the Hierarchical Task Analysis (HTA) and task network allows analysts to gain a deep understanding of the system under analysis [17].

STPA identifies potential failures in control and feedback mechanisms within a system to develop effective risk controls [18], and highlights the importance of integrating safety controls into the design and operational phases [6], [19], [20], [21], [22].

On the other hand, Bayesian Networks (BNs) provide a probabilistic approach to risk assessment, modelling uncertainties and interdependencies among risk factors. They have been used to evaluate collision risks and the resilience of autonomous systems, providing insights into potential failure points and mitigation strategies [23], [24], [25], [26], [27].

Finally, the majority of the risk analysis approaches that are being researched for MASS are hybrid methods as they combine system-based hazard identification methods with traditional risk assessment methods like STPA with BN [28]. The remaining risk assessment methods for MASS involves methods/frameworks that approach the issue innovatively like the comprehensive combined methodology proposed by DNV where the risk assessment is based on the below novel equation [8], [29]:

Risk = Mitigation measure effectiveness · Severity of outcome from Risk Control Measures (RCM) failure.

In order to assess the final risk, the method includes three initial stages [30], [31]:

- Hazard Identification (HAZID)
- Fault Tree Analysis (FTA)
- Risk Control Options (RCO) and RCM

3. Methodology

The methodology used in the present study relies on the combination of two separate methods. The primary methodology used is NET-HARMS, which is a method that is being applied for the first time in MASS. Then, the Risk Matrix from RBAT, recommended by DNV, is used to address some gaps in the probability assessment stage of NET-HARMS. The probability in NET-HARMS is calculated through the risk occurrence frequency and the probability of emergent risks is often rated as high as the related task risk is viewed as having occurred [7]. However, in our case this is not feasible as not enough risk occurrence data are available.

3.1 Net-HARMS

Net-HARMS [32] is a risk analysis methodology grounded in systems theory, integrating HTA [33], Event Analysis of Systemic Teamwork (EAST) [16], and SHERPA [34] principles. It anticipates hazards in complex sociotechnical settings by meticulously describing systems and applying a specialized taxonomy to pinpoint both tasks and emergent risks. Arising from a critique of current risk assessment tools which often neglect modern theories on how accidents occur and fail to recognize systemic workplace hazards, Net-HARMS stands out by its systemic risk identification capabilities and its focus on emergent risks resulting from interactions within the system [32].

Net-HARMS is the newest systems-based risk assessment method in Human Factors Engineering (HFE), so it has been applied to only three case studies. Firstly, it had been applied to the led outdoor activity sector [32] and it was recently used to identify task and emergent risks throughout the design lifecycle of railway level crossings [7]. Additionally, it was employed in the context of elite sports to pinpoint potential risks that could jeopardize the performance of a cycling team [35]. Nevertheless, the approach is inherently versatile and can be utilized for assessing risks across any field.

Net-HARMS offers two main improvements compared to existing risk assessment methods: it allows for a comprehensive system-wide risk detection, as opposed to focusing solely on immediate 'sharp-end' risks, and it uncovers 'emergent risks' spawned by the confluence of various factors. Designed for simplicity, the method involves crafting an HTA that maps out the system in question. This HTA then transforms into a task network that outlines crucial tasks for achievement of overall goal, as well as how they interrelate.

The method proceeds with the Net-HARMS risk mode taxonomy, applied to each node of the task network to find potential task risks and their consequences. The taxonomy's second application seeks out emergent risks, which could manifest when task-related risks have a ripple effect on other areas of work. This stage is crucial, as it helps predict and address novel and unforeseen risks linked to subpar performance within the system. Analysts then assess the likelihood and severity of these risks categorized as low, medium, or high. In this study this step is performed through RBAT methodology.

3.2 RBAT

A method tailored for assessing risks linked with MASS concepts has been devised. Many of the identified risks for MASS are anticipated to involve control issues, primarily originating from software failures. Due to the inherent difficulty in predicting the probability of such risks, it was decided to depart from the traditional risk definition based on probability and consequence. Similar challenges have been encountered in other sectors, like the automotive industry, in their safety assurance endeavors [8]. While drawing on these experiences, the proposed approach seeks to adapt them to suit the specific requirements of the maritime industry without straying from established practices and frameworks.

Rather than adhering strictly to the conventional risk definition (probability * consequence), the RBAT method evaluates risk by considering:

- The severity of the worst-case outcome resulting from an undesired event (**Table 1**)
- The effectiveness of the concept's measures in preventing losses (**Table 2**).

Risk Acceptance Criteria (RAC) have been suggested to enable demonstrating risks being reduced to As Low As Reasonably Practicable (ALARP). These criteria have been compared and aligned with other RAC commonly found in safety standards, including those utilized by different industries. The two indexes which together form the risk matrix used in RBAT are presented below. Criteria for what is considered acceptable and unacceptable risk levels are also suggested. **Table 3** merges the severity index (**Table 1**) and mitigation effectiveness index (**Table 2**) to form the proposed risk matrix for RBAT, accompanied by risk acceptance criteria. It suggests applying the ALARP principle for risk evaluation:

- High (red region): Risk must be reduced regardless of costs.
- Medium (yellow ALARP region): Risk should be reduced to a level that is reasonably practicable.
- Low (green region): Risk is negligible, and no reduction is necessary.

Table 1. *Severity Index for worst-case outcomes.*

Severity	Effects on human safety
No effect	No injuries
Negligible	Superficial injury
Minor	Single injury or multiple minor injuries
Significant	Single serious or multiple injuries
Severe	Single fatality or multiple serious injuries
Catastrophic	Multiple fatalities (more than one)

Table 2. *Effectiveness of mitigation layers.*

Effectiveness	Description
Very high	At least three effective independent mitigation layers that for the assessed scenario can prevent losses regardless failure cause.
High	At least two effective independent mitigation layers that for the assessed scenario can prevent losses regardless failure cause.
Medium	At least one effective independent mitigation layer that for the assessed scenario can prevent losses regardless failure cause.
Moderate	At least one internal mitigation layer that can prevent losses from random hardware failures. The control function has additional capacities for self-recovery from other types of failures, however, for the assessed scenario these are not effective regardless failure cause.
Low	The control function has some capacities for self-recovery, however for the assessed scenario these are expected to have a limited effect.

4. Results

The above methodology will be applied to the autonomous tugboat concept developed in the MOSES project. This concept involves autonomous tugboats operated as a swarm (cooperatively) in order to support the manoeuvring and the docking/undocking processes of large containerships at their arrival/departure to/from the Deep Sea Shipping (DSS) port [36]. The control architecture of the autonomous tugboat swarm consists of the following components:

i) detection module, which is responsible for sensor data-processing
ii) path planning module, which is responsible for autonomous navigation and manoeuvring and includes all the motion control operations and
iii) control module, which translates the high-level decisions from the navigation algorithms into actionable steering and propulsion commands.

Table 3. *Risk matrix based on evaluation of available risk mitigating measures.*

Effectiveness of risk mitigation layers	Severity					
	No effect	Negligible	Minor	Significant	Severe	Catastrophic
Low	Low	Medium	High	High	High	High
Moderate	Low	Low	Medium	High	High	High
Medium	Low	Low	Medium	Medium	High	High
High	Low	Low	Low	Medium	Medium	High
Very high	Low	Low	Low	Low	Medium	Medium
Extremely high	Low	Low	Low	Low	Low	Medium

MOSES targets various levels of autonomy for the tugboats, where monitoring and control functionalities will be allocated to humans onboard the tugboats, the AI algorithms, and the remote operator in the Shore Tugboat Control Station (STCS).

With respect to the levels of autonomy proposed in the DNV Class Guidelines for Autonomous and remotely operated ships, the autonomous tugboat swarm will operate in the following modes in terms of autonomy level:

- Manual navigation with decision support by the remote operator in the STCS (Decision supported function).
- Autonomous swarm operation with remote control capability by the remote operator in the STCS and manual override capability from the tugboat Captain, which can be conducted at any time (Self-controlled function, human-in-the-loop).

The first step in the application of Net-HARMS involves the definition of the overall goal, which is : Vessel's manoeuvring and docking/undocking. The next step involves decompossing the overall goal into sub-goals and then each sub-goal into more specific sub-goals and operations, until the necessary level of detail is attained, so all the goals will be clearly defined and actionable. The final step, in order to create the HTA diagram (**Figure 1**), is to determine the plans (i.e., linear: Do 1, then do 2, then do 3, then EXIT, selection: Do 1, then do 2 or 3 as required) that specify the sequence of goals, sub-goals, and operations.

After finalizing the HTA, a task network is constructed to identify tasks and potential risks. Task networks, composed of nodes (tasks) and arrows (relationships), represent HTA outputs and show task interactions. Therefore, we can understand work system coupling using task networks. Relationships between tasks are included if they are sequential, concurrent, influencing, or dependent on each other.

Journal of Physics: Conference Series **2867** (2024) 012046 doi:10.1088/1742-6596/2867/1/012046

In our research, the task network was constructed by taking the sub-goals that are directly related to the autonomous process as well as the sub-goals that refer to the adequacy of human element and control systems (see 1. Idle Situation and 4. Manoeuvring to/away from dock, (**Figure 1**). Relationships between tasks are included if they are sequential, concurrent, influencing, or dependent on each other. However, in this paper we will indicatively present the results for the sub-goal: 1.2.1 Check functionality of AP (**Figure 2**), which is related with 7 sub-goals within the task network. On the left side of **Figure 2**, a section of the entire task network is presented and on the right side, the sub-goal "Check functionality of AP" is highlighted, along with the sub-goals linked to this sub-goal.

Then, utilizing the Net-HARMS taxonomy and the task network, 'task risks' during each task are identified. Task risks may occur if tasks are not performed optimally. To identify task risks, consider 10 sub-optimal ways using the Net-HARMS risk mode taxonomy and domain expertise for each task step. The taxonomy covers task performance aspects (i.e., task omitted or task completed inadequately), communication between agents (i.e., information not communicated or wrong information communicated), and environment. We ought to analyse each risk mode methodically for every task to determine potential risks. An extract of the vessel's manoeuvring and docking/undocking task risks is presented in **Table 4**.

Then, the Net-HARMS taxonomy collaborates with the task network to identify emergent risks that could occur as a result of the interaction of task risks. This involved examining each set of related tasks in the task network and identifying what the impact of task risks would be on related tasks. For each credible emergent risk, we recorded a description of the risk, its consequences, and provided a rating of the severity index for worst-case consequence, the effectiveness index of mitigation layers, and the risk index. An extract of the vessel's manoeuvring and docking/undocking emergent risks is presented in **Table 5**. In this table are presented all the related, with the task 1.2.1, tasks, which could be affected if the task 1.2.1 ommited. In the next columns, each risk is categorized by its emergent risk mode, description, and consequences, and is assessed in terms of severity of worst-case concequense, mitigation layers effectiveness, and overall risk level. As independent mitigation layers within our system there are the STCS, who can monitor and remote control the tugboats and the Captains, who can also monitor and take control of the tugboats.

The application of NET-HARMS to the autonomous tugboat system revealed several critical task risks associated with various sub-goals. Firstly, the adequacy, training level, health, and readiness of human in the loop along with safeguards in order to check them (Sub-goal 1.1), are crucial as the inadequate or incomplete check of them could compromise operational safety. Similarly, failing to thoroughly assess the functionality of key system components such as the Autopilot (AP) and Data Acquisition (DAQ) systems (Sub-goals 1.2.1 and 1.2.2) could result in undetected malfunctions or damage, which might not only disrupt the autonomous maneuvering process but also pose significant hazards if incorrect or insufficient information is communicated regarding these systems' operational status. Moreover, the analysis highlighted substantial risks linked to the training and use of AI algorithms (Sub-goal 1.4). The use of incorrect or inadequate data for training AI algorithms could lead to suboptimal path planning and collision avoidance, potentially resulting in collisions with static or dynamic obstacles. The study also underscored the importance of timely and accurate communication among crew members and system controllers, as delays or errors in relaying mission-critical information (Sub-goals 3.9 and 4.7) could prevent the successful transition to subsequent phases of autonomous operation, thereby elevating the risk of operational failures and safety incidents.

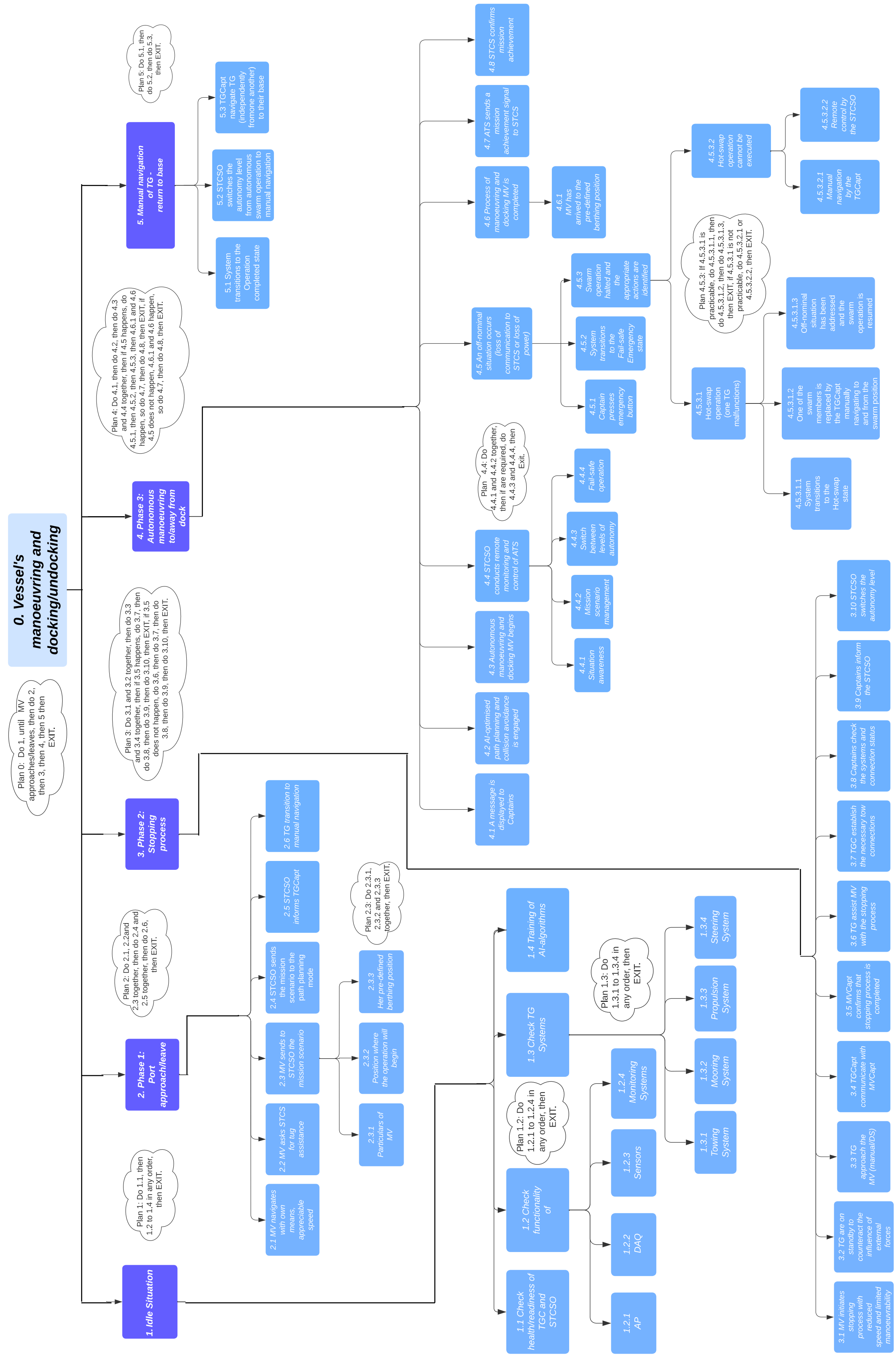

Figure 1. *HTA diagram.*

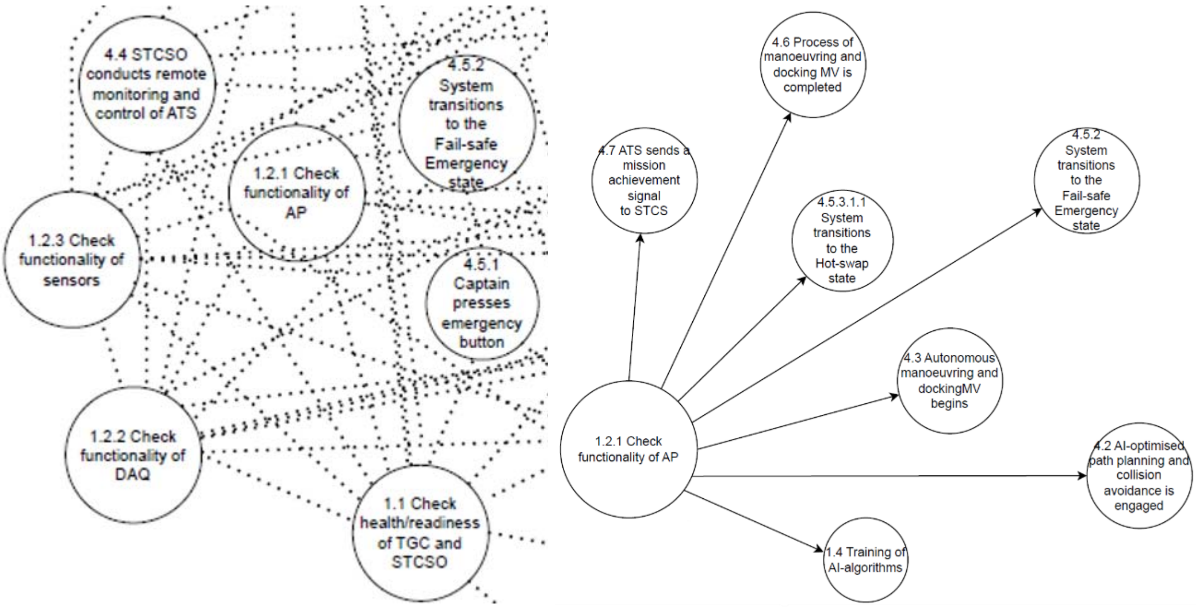

Figure 2. *Illustrative example of the sub-goal 'Check functionality of AP'.*

Table 4. *Extract of the vessel's manoeuvring and docking/undocking task risks.*

HTA Sub-goal	Risk Mode	Task Risk Description	Task Risk Consequences
1.1 Check adequacy/health /readiness of TGC and STCSO	T1	The check of the training-familiarization level of crew/captains/STCSO was done before the end of the training induction.	It is not guaranteed that TGC and STCSO are adequately trained.
	T2	The check of the training-familiarization level of crew/captains/STCSO was not done.	It is not guaranteed that TGC and STCSO are trained.
	T2	The check of TGC was not done.	Some of the members may not be healthy/ready for the process.
	T2	The check of STCSO was not done.	STCSO may not be healthy/ready for the process.
	T3	The check of the training-familiarization level of crew/captains/STCSO was inadequately done.	It is not guaranteed that TGC and STCSO are adequately trained.
	T3	The check of TGC was done only once.	TGC may face issues after the check.
	T3	The check of STCSO was done only once.	STCSO may face issues after the check.
4.2 AI-optimised path planning and collision avoidance is engaged	T1	AI-optimised path planning and collision avoidance is engaged too late.	Potential for incorrect or not optimised path planning and implementation. Potential for collision with static or dynamic obstacles.
	T2	AI-optimised path planning and collision avoidance is not engaged.	
4.4 STCSO conducts remote monitoring and control of ATS	T1	Remote monitoring and control of ATS delayed.	STCSO has not the situation awareness. STCSO does not manage mission scenario. STCSO is not ready to switch between levels of autonomy. STCSO is not ready for a fail-safe operation.
	T2	STCSO does not conduct remote monitoring and control of ATS.	
	T3	STCSO conducts remote monitoring and control of ATS inappropriately.	
	T4	MS are inadequate.	

The results from the application of the risk matrix of RBAT to the identified emergent risks linked with the sub-goal 1.2.1 Check functionality of AP, revealed 7 Medium risks, 4 High risks, and 2 Low risks. In general, the risks tended themselves in two different areas, either on the Severe/High area or the Severe-Catastrophic/Medium area. The second area, which is high-risk and needs our attention to reduce the risk, has arisen due to the failure of one of the existing mitigation layers. However, considering that the probability of the AP failing, and this failure potentially causing the system not to transition to the Fail-safe Emergency state, is very low, the actual risk is lower than the estimated risk. Therefore, it is not necessary to implement additional risk control measures. However, the Fail-Safe Emergency State would constitute an additional independent mitigation layer if it was activated autonomously and did not require the captain to press a button to activate it.

Table 5. *Extract of the vessel's manoeuvring and docking/undocking emergent risks.*

Related Task	Emer. Risk Mode	Emergent Risk Description	Emergent Risk Consequences	Emer. Risk Severity	Emer. Risk Mit. Eff.	Risk Level
4.2 AI-optimised path planning and collision avoidance is engaged	T1	AI-optimised path planning and collision avoidance is engaged too late.	Potential for incorrect or not optimised path planning and implementation. Potential for collision with static or dynamic obstacles.	Severe	High	Medium
4.3 Autonomous manoeuvring and docking MV begins	T1	The beginning of the autonomous manoeuvring and docking has been delayed.	Potential traffic caused by MV's delayed docking/undocking.	Minor	High	Low
4.5.2 System transitions to the Fail-safe Emergency state	T1	System's transitioning to the Fail-safe Emergency state delayed.	Swarm operation does not halt immediately and the appropriate actions are not identified.	Severe	Medium	High
4.5.3.1.1 System transitions to the Hot-swap operation	T1	System's transitioning to the Hot-swap operation delayed.	The swarm member is not replaced by the TGCapt manually navigating to/from the swarm position.	Severe	Medium	High
4.6 Process of manoeuvring and docking MV is completed	T1	Completion of the manoeuvring and docking MV delayed.	Arrival to the pre-defined berthing position delayed. Potential traffic caused by MV's delayed docking/undocking.	Minor	High	Low
4.7 ATS sends a mission achievement signal to STCSO	T2 / C1	ATS does not send a mission achievement signal to STCSO.	STCSO cannot confirm mission achievement. So, the System cannot transition to the Operation completed state.	Significant	High	Medium
1.4 Training of AI-algorithms	T3	Use of incorrect data to train AI-algorithms. Use of inadequate data to train AI-algorithms. Inadequate testing.	Potential for incorrect or not optimised path planning and implementation. Potential for collision with static or dynamic obstacles. TGS may not react to changes of the MV's operational parameters. TGS may not comply with port navigational restrictions.	Catastrophic	High	High

5. Conclusions

The application of the NET-HARMS methodology to MASS has demonstrated significant advantages in hazard identification and risk assessment. NET-HARMS is more suitable for MASS systems, which are at the design stage, as it can comprehensively cover risks across the entire system, including government and regulatory levels [7]. Therefore, the application of NET-HARMS can assist in the design and construction of remote-control centres by adopting human-centred design principles and HFE techniques and standards. In addition, considering that MOSES project operates in different levels of autonomy, by the application of NET-HARMS to MOSES, we face the problem referred in a previous paper [15], concerning the assessment of how safety can be affected according to the changes on the level of autonomy and the role of human.

The method's structured and systematic approach makes it easy to learn and apply, requiring minimal training. Its generic nature allows NET-HARMS to be used across various domains. However, in newly developed areas such as MASS, the final steps involving probability calculation cannot be conducted due to insufficient data, necessitating the combination of NET-HARMS with other risk assessment methods.

Despite its benefits, the NET-HARMS analysis, especially for complex systems such as MASS, can be time-consuming [32], as it largely dependent on the number of tasks and the relationships between them within the task network. Moreover, it was observed high levels of repetition, with risks often being identified multiple times.

As one of the first applications in this domain, more complex applications must be studied to assess and compare the emerging risks in order to make NET-HARMS a valuable tool in enhancing safety and operational reliability. Since conventional and autonomous ships will coexist for a considerable amount of time, a more complex future research may be the identification of possible risks associated with both completely autonomous and remotely operated ships [6].

Acknowledgments

MOSES project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 861678. The content of this document reflects only the authors' view and the Agency is not responsible for any use that may be made of the information it contains.

References

[1] M. Wahlström, J. Hakulinen, H. Karvonen, and I. Lindborg, "Human Factors Challenges in Unmanned Ship Operations – Insights from Other Domains," *Procedia Manuf.*, vol. 3, pp. 1038–1045, Jan. 2015, doi: 10.1016/j.promfg.2015.07.167.

[2] C. Fan, K. Wróbel, J. Montewka, M. Gil, C. Wan, and D. Zhang, "A framework to identify factors influencing navigational risk for Maritime Autonomous Surface Ships," *Ocean Eng.*, vol. 202, p. 107188, Apr. 2020, doi: 10.1016/j.oceaneng.2020.107188.

[3] E. Veitch and O. Andreas Alsos, "A systematic review of human-AI interaction in autonomous ship systems," *Saf. Sci.*, vol. 152, p. 105778, Aug. 2022, doi: 10.1016/j.ssci.2022.105778.

[4] X.-Y. Zhou, Z.-J. Liu, F.-W. Wang, and Z.-L. Wu, "A system-theoretic approach to safety and security co-analysis of autonomous ships," *Ocean Eng.*, vol. 222, p. 108569, Feb. 2021, doi: 10.1016/j.oceaneng.2021.108569.

[5] M. Abilio Ramos, I. B. Utne, and A. Mosleh, "Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events," *Saf. Sci.*, vol. 116, pp. 33–44, Jul. 2019, doi: 10.1016/j.ssci.2019.02.038.

[6] X.-Y. Zhou, Z.-J. Liu, F.-W. Wang, Z.-L. Wu, and R.-D. Cui, "Towards applicability evaluation of hazard analysis methods for autonomous ships," *Ocean Eng.*, vol. 214, p. 107773, Oct. 2020, doi: 10.1016/j.oceaneng.2020.107773.

[7] P. M. Salmon *et al.*, *Handbook of Systems Thinking Methods*, 1st ed. Boca Raton: CRC Press, 2022. doi: 10.1201/9780429281624.

[8] DNV, "Specific MASS Concepts and Risk Evaluation Techniq 3." 2021.

[9] Ø. Rødseth and H. Nordahl, *Definition of autonomy levels for merchant ships, Report from NFAS, Norwegian Forum for Autonomous Ships, 2017-08-04*. 2018. doi: 10.13140/RG.2.2.21069.08163.

[10] Allianz, "Safety and Shipping Review," 2017.
[11] DNV, "Autonomous and remotely operated ships." DNV, 2018.
[12] M. Kim, T.-H. Joung, B. Jeong, and H.-S. Park, "Autonomous shipping and its impact on regulations, technologies, and industries," *J. Int. Marit. Saf. Environ. Aff. Shipp.*, vol. 4, no. 2, pp. 17–25, Apr. 2020, doi: 10.1080/25725084.2020.1779427.
[13] Koji Wariishi, "MARITIME AUTONOMOUS SURFACE SHIPS : DEVELOPMENT TRENDS AND PROSPECTS — HOW DIGITALIZATION DRIVIES CHANGES IN MARITIME INDUSTRY—." Mitsui & Co., 2019.
[14] Bureau Veritas, "Guidelines for Autonomous Ships." 2019.
[15] N. P. Ventikos, P. Siokouros, and A. Koimtzoglou, "Application of the EAST-BL method on a MASS system for Hazard Identification and Risk Assessment," *J. Phys. Conf. Ser.*, vol. 2618, no. 1, p. 012014, Oct. 2023, doi: 10.1088/1742-6596/2618/1/012014.
[16] Stanton, N.A. & Harvey, C., "Beyond human error taxonomies in assessment of risk in sociotechnical systems: A new paradigm with the EAST 'Broken-Links' approach. Ergonomics, 60, 2." 2017.
[17] A. Hulme *et al.*, "Systems thinking-based risk assessment methods applied to sports performance: A comparison of STPA, EAST-BL, and Net-HARMS in the context of elite women's road cycling," *Appl. Ergon.*, vol. 91, p. 103297, Feb. 2021, doi: 10.1016/j.apergo.2020.103297.
[18] Leveson & Thomas, "STPA Handbook," 2018.
[19] M. Chaal, O. A. Valdez Banda, J. A. Glomsrud, S. Basnet, S. Hirdaris, and P. Kujala, "A framework to model the STPA hierarchical control structure of an autonomous ship," *Saf. Sci.*, vol. 132, p. 104939, Dec. 2020, doi: 10.1016/j.ssci.2020.104939.
[20] O. A. Valdez Banda, S. Kannos, F. Goerlandt, P. H. A. J. M. Van Gelder, M. Bergström, and P. Kujala, "A systemic hazard analysis and management process for the concept design phase of an autonomous vessel," *Reliab. Eng. Syst. Saf.*, vol. 191, p. 106584, Nov. 2019, doi: 10.1016/j.ress.2019.106584.
[21] N. P. Ventikos, A. Chmurski, and K. Louzis, "A systems-based application for autonomous vessels safety: Hazard identification as a function of increasing autonomy levels," *Saf. Sci.*, vol. 131, p. 104919, Nov. 2020, doi: 10.1016/j.ssci.2020.104919.
[22] ITOH, YUZUI, SHIOKARI, ISHIMURA, MIYAKE, and KUDO, "Risk Assessment of Autonomous Ship Systems." 2021.
[23] C. A. Thieme, I. B. Utne, and S. Haugen, "Assessing ship risk model applicability to Marine Autonomous Surface Ships," *Ocean Eng.*, vol. 165, pp. 140–154, Oct. 2018, doi: 10.1016/j.oceaneng.2018.07.040.
[24] J. Zhang, M. Jin, C. Wan, Z. Dong, and X. Wu, "A Bayesian network-based model for risk modeling and scenario deduction of collision accidents of inland intelligent ships," *Reliab. Eng. Syst. Saf.*, vol. 243, p. 109816, Mar. 2024, doi: 10.1016/j.ress.2023.109816.
[25] C. Katrakazas, M. Quddus, and W.-H. Chen, "A new integrated collision risk assessment methodology for autonomous vehicles," *Accid. Anal. Prev.*, vol. 127, pp. 61–79, Jun. 2019, doi: 10.1016/j.aap.2019.01.029.
[26] C.-C. Chou, C.-N. Wang, and H.-P. Hsu, "A novel quantitative and qualitative model for forecasting the navigational risks of Maritime Autonomous Surface Ships," *Ocean Eng.*, vol. 248, p. 110852, Mar. 2022, doi: 10.1016/j.oceaneng.2022.110852.
[27] M. M. Abaei, R. Hekkenberg, A. BahooToroody, O. V. Banda, and P. Van Gelder, "A probabilistic model to evaluate the resilience of unattended machinery plants in autonomous ships," *Reliab. Eng. Syst. Saf.*, vol. 219, p. 108176, Mar. 2022, doi: 10.1016/j.ress.2021.108176.
[28] I. B. Utne, B. Rokseth, A. J. Sørensen, and J. E. Vinnem, "Towards supervisory risk control of autonomous ships," *Reliab. Eng. Syst. Saf.*, vol. 196, p. 106757, Apr. 2020, doi: 10.1016/j.ress.2019.106757.
[29] DNV, "Testing on Specific Cases of MASS Connccepts 4." 2021.
[30] DNV, "Study of the risks and regulatory issues of specific cases of MASS 2." 2020.
[31] DNV, "Study of the risks and regulatory issues of specific cases of MASS 3." 2020.
[32] C. Dallat, P. M. Salmon, and N. Goode, "Identifying risks and emergent risks across sociotechnical systems: the NETworked hazard analysis and risk management system (NET-HARMS)," *Theor. Issues Ergon. Sci.*, vol. 19, no. 4, pp. 456–482, Jul. 2018, doi: 10.1080/1463922X.2017.1381197.
[33] N. A. Stanton, "Hierarchical task analysis: Developments, applications, and extensions," *Appl. Ergon.*, vol. 37, no. 1, pp. 55–79, Jan. 2006, doi: 10.1016/j.apergo.2005.06.003.
[34] Embrey, "SHERPA: A systematic human error reduction and prediction approach." 1986.
[35] A. Hulme, N. A. Stanton, G. H. Walker, P. Waterson, and P. M. Salmon, "Testing the Reliability and Validity of Net-HARMS: A New Systems-Based Risk Assessment Method in HFE," in *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, vol. 219, N. L. Black, W. P. Neumann, and I. Noy, Eds., in Lecture Notes in Networks and Systems, vol. 219. , Cham: Springer International Publishing, 2021, pp. 354–362. doi: 10.1007/978-3-030-74602-5_51.
[36] Ventikos NP, Louzis K, Themelis N, et al., "MOSES Autonomous tugboat swarm operation: Operational scenarios, requirements, and architecture.," 2022.