Towards Simulation-based Verification of Autonomous Navigation Systems

Tom Arne Pedersen[*], **Jon Arne Glomsrud** [†] **and Odd Ivar Haugen**[‡]
Digital Assurance Program, Group Technology and Research, DNV GL, Norway

ABSTRACT

Autonomous ships are expected to change water-based transport of both cargo and people, and large investments are being made internationally. There are many reasons for such transformation and interest, including shifting transport of goods from road to sea, reducing ship manning costs, reduced dangerous exposure for crew, and reduced environmental impact.

Situational awareness (SA) systems and Autonomous navigation systems (ANS) are key elements of autonomous ships. Safe deployment of ANS will not be feasible based on real-life testing only. The assurance of autonomous ships and systems will require large-scale, systematic simulation-based testing in addition to assurance of the development process.

DNV GL proposes to use a digital twin, that is a digital representation of key elements of the autonomous vessel as a key tool for the simulation-based testing, focusing on functional testing, failure tolerance, and performance aspects. The digital twin contains comprehensive mathematical models of the ship and its equipment, including all sensors and actuators. The complete simulation-based test system complementing the digital twin should consist of a virtual world to simulate environmental conditions, geographical information and interaction with other maritime traffic and obstacles. Finally, the test system must include a test management system that controls the simulations in the digital twin and the virtual world, generates test scenarios as well as evaluates the test scenario results. The scenario generation should automatically search for low system performance, and ultimately establish sufficient coverage of the possible scenario space. The test scenario evaluation should automatically consider safety, conformance to collision regulations at sea (COLREGs), and possibly the efficiency of the ship navigation.

This paper presents a comprehensive prototype of a test system for ANS. Key topics will be simulation-based testing, interfacing between the simulator and ANS, cooperation with ANS manufacturers, dynamic test scenario generation and automatic assessment towards COLREGs.

Keywords: Autonomous navigation, digital twin, simulation-based testing, dynamic test scenario, automatic test scenario generation

1 INTRODUCTION

Ships have always been operated by seafarers. The crew size has depended upon the size and type of ship, and the type of mission. In recent years, substantial development has been achieved in sensor technology, machine learning, automation and connectivity. This means that, at least in theory, it is possible to reduce or even remove the crew from the ship. However, this will require either shore-based remotely monitored and operated ship systems, or the ship systems operating autonomous based on algorithms.

Remotely controlled or autonomous functions are not necessarily implemented only to reduce cost, but also because of safety reasons. According to (Safety4sea, 2018), about 80% of marine accidents are caused by human errors. Working conditions for the crew and as well as lower emissions are also important factors for this shift. The possibility of using autonomous and remotely operated vessels are also introducing novel or changed transport systems and business models

---

[*] Corresponding author: +47 95280695, tom.arne.pedersen@dnvgl.com
[†] Corresponding author: +47 92403158, jon.arne.glomsrud@dnvgl.com
[‡] Corresponding author: +47 91715040, odd.ivar.haugen@dnvgl.com

where e.g. smaller unmanned vessels can be used the last mile bringing cargo from a mother ship to smaller less area-demanding harbours.

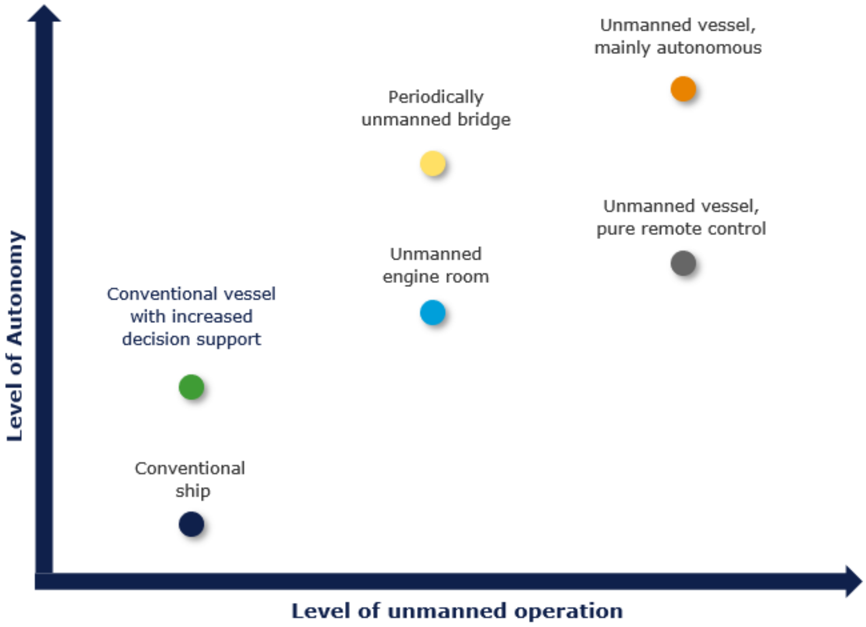


Figure 1: Level of Autonomy vs level of unmanned operation

In the maritime industry, autonomous ships are on everyone's lips, but what this actually entails can vary widely. Several definitions of the level of autonomy exists, and what is common is to define the level of autonomy as a system's increasing ability to operate without human control or intervention. The scale ranges from no autonomy where the human operator needs to take all decisions, to fully autonomous without a human operator in the loop. Autonomous does not equal unmanned and many levels of autonomy does not contain this aspect. Figure 1 map different levels of autonomy (vertical axis) against level of unmanned operation (horizontal axis). Conventional ships are placed in the lower left corner, with a low level of autonomy and unmanned operation. Ships with added decision support have a higher level of autonomy and are thus placed higher on the left part of the figure, though still with a low level of unmanned operation. Presently, the engine rooms onboard ships are unmanned certain time periods, and it is required that the engine room can operate at least 24 hours without manual monitoring and control. The engine room operator, however, must be onboard the ship. Unmanned engine rooms, found in the center of Figure 1, indicates that the engine room can operate without manual control from onboard crew for weeks or months, and that the control and monitoring is done from an on-shore control site. A periodically unmanned bridge is also placed in the middle of Figure 1. At the right part of Figure 1, unmanned ships that are either remotely controlled or autonomous are found. One may notice that a typical vessel is not either conventional or autonomous/remote operated, but instead some ship systems may be unmanned, while others are not.

To navigate safely, the ship crew or navigation system needs to detect any elements that may affect the planned path of the ship. In Figure 2, the ship navigation function is broken down into sub tasks.

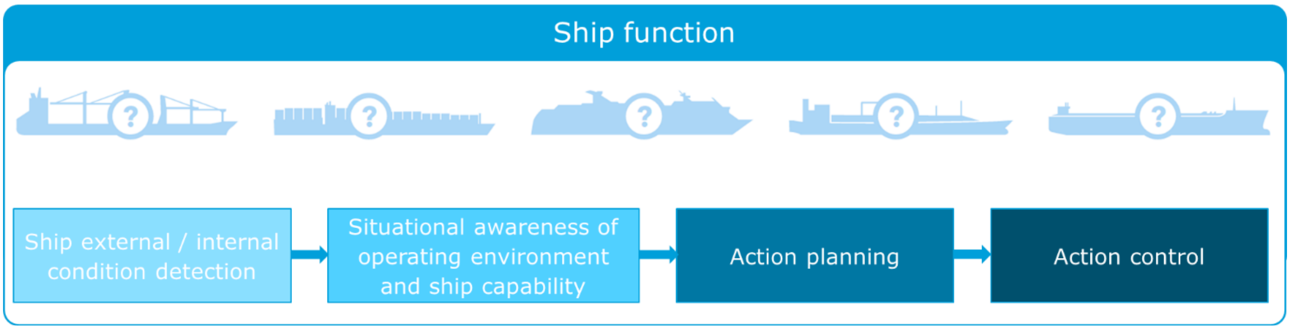

Ship function
Ship external / internal condition detection
Situational awareness of operating environment and ship capability
Action planning
Action control

Figure 2: Ship functions broken down into sub tasks. Based on (Vartdal, Skjong, & St. Clair, 2018)

Initially, the ship navigator needs to know the external and internal operational and ship conditions, including geography, bathymetry, fixed or floating objects, and weather conditions together with the conditions of the ship's equipment. A priori information may come from e.g. Electronic Navigation Charts (ENC), Automatic Identification Signatures (AIS), etc. Not all ships transmit AIS data and not all AIS data are reliable, thus exteroceptive sensors need to be used in addition to be able to detect all objects relevant for the navigation. Examples of exteroceptive sensors are radar, camera, infrared camera and lidar.

To achieve situational awareness, the different elements need to be classified and their states determined. Computer vision using camera is a field that has come far when it comes to detecting and classifying surrounding objects, but in the maritime industry there is still a long way to go. Computer vision is usually based on machine learning, and machine learning needs to be trained using pre-existing pictures or video footage which are currently limited in the maritime industry.

Once the system has analyzed the situation and sufficient situational awareness is achieved, the course of action needs to be planned. The planning is done by the ANS using information from the predefined ship mission and the predefined set of navigation rules such as the COLREG. COLREG is written for human navigators and since many of the requirements are qualitative and open for interpretation and situational judgement to cover as many different scenarios as possible, it is difficult to develop an ANS based on this.

The last sub task is action control. The engines and rudder, or thrusters are operated to navigate the ship.

Risk is an important factor to consider while navigating. An autonomous navigation system will always need to evaluate its performance, and if the performance it not within acceptable limits, or the risk of continue its ongoing operation is considered too high, the vessel should enter a defined minimum risk condition (MRC). The MRC will vary dependent on location, operation, surroundings etc. and the resulting action may be for example to stop and go into DP mode, to go to nearest port or similar.

When introducing new technology or using existing technology for new purposes, uncertainties are also introduced. The risks of safe operation in unmanned shipping have among other been studied in the MUNIN project (Rødseth & Burmeister, 2015). These risks need to be adequately handled, and in (Heikkilä, Tuominen, Tiusanen, Montewka, & Kujala, 2017), the safety qualification process is solved using a goal-based safety case approach. This process is based on the recommended practice for technical qualification, DNVGL RP-A203 (DNVGL, 2017). During a qualification process, the safety goals and risks are identified, and qualification activities are then performed to collect evidence of reaching the goals and mitigating the risks.

For the perspective of assurance and testing, it is of utmost importance to ensure that the ANS algorithms are safe and do not cause accidents. That is, the ANS should go through a qualification process. Testing of the actual ANS will be an important activity in providing evidence that the ANS is safe. Testing may be done in real life using the actual ship, in the virtual world using simulators, or in a combination. Real-life testing is too time consuming and many required test scenarios will be impossible to test, thus a combination of simulation-based and real-life testing would be the preferred solution. The real-life testing could be used to gather knowledge and construct scenarios for simulation-based testing, and to produce data to validate the digital twins, digital models and simulators (Wood, et al., 2019).

In the next sections we explore simulation-based verification, unpack the components of a test verification system, the role of an open simulation platform as an important enabler, and discuss the evolution of test scenarios.

2 SIMULATION-BASED TESTING

Simulation-based testing will be an important tool when collecting evidence of safe ANS algorithms. A proposal for a test system is shown in the Figure 3 consisting of the test management system and a virtual world. The different parts of the test system are explained in the following.

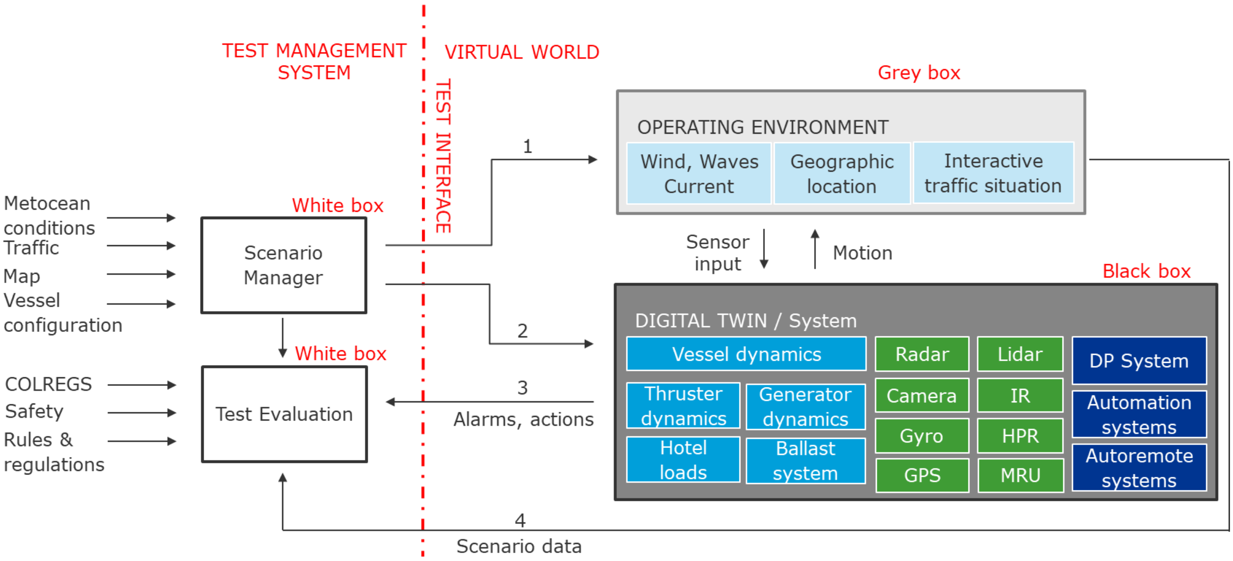


Figure 3: Test system for autonomous navigation systems

2.1 Digital twin

The digital twin is a vital part of the test system shown in Figure 3. The digital twin is a virtual representation of a particular ship, called *own ship*, that will be controlled by the ANS under test. It is a comprehensive mathematical model of the real ship and includes models of the ship-specific hull dynamics including fluid/hydrodynamics, its power system, propulsion system, ballast system, sensors and actuators etc, in addition to emulated control system hardware running actual control system software. Control system software included in the digital twin may be dynamic positioning (DP) system, power management system (PMS), automation control system etc (see the lower right box in Figure 3). The different models need to be accurate enough to capture relevant dynamics of the ship, and the control system should “believe” it is controlling the actual ship systems and not the simulated ship systems. Necessary control system software to include in the test setup depends on the algorithms that are tested.

2.2 Operating environment

The operating environment is another vital part of the virtual world, see Figure 3. To play out relevant and realistic test scenarios, it is important to have full control of the environmental conditions such as wind, waves and current, in addition to geographic location and interactive traffic, e.g. other ships called *target ships*. The word *interactive* is in this context important. If using e.g. historic AIS data recorded from ships in a specific area as basis for simulating the target ships, these will not be interacting with the own ship, but only replaying the recorded AIS information. Instead, AIS data could be used as input to construct test scenarios, and when simulating the test scenario, the target ships can interact both with each other and with own ship in the same way as ships interact in real life. To achieve this, also the target ships need to be navigated, either by a human navigator or by other ANS algorithms in the simulation. The test system should therefore include various ANS to ensure that the own ship ANS is robust towards a variety of target ship behaviours. Occasionally, other ships may not behave as expected and this also needs to be handled by an ANS, thus the operating environment will also include target ships not behaving in full compliance with COLREGs.

2.3 Test management system

The test management system shown in Figure 3 consists of two parts. The scenario manager uses environmental conditions, traffic, location and vessel configuration as input for setting up scenarios used for testing the ANS. The testing should focus on operational and failure scenarios, but in addition, performance testing is also possible in the simulation-based testing given the digital twin has the sufficient accuracy.

The second part of the test management system is the test evaluation. Using results from the simulation of the test scenarios, the ANS algorithms will be evaluated against COLREGs, safety and other relevant rules and regulations. Test scenario evaluation is treated more thoroughly in chapter 3.

2.4 Test interface

From Figure 3, one may notice the test interface between the test management system and the virtual world. Looking at arrow 1, it is important that the scenario manager has full control of the operating environment, configuring the test scenario exactly as intended. It must be possible to initiate position, course and speed of the target ships and in addition be able to set path plans or waypoints for the target ships and decide to which degree they shall follow COLREGs. Environmental disturbances and location are other elements that are important for the scenario manager to control.

The scenario manager also needs to interface the own ship controlled by the ANS algorithm under test, see arrow 2 in Figure 3. Initial position, course and speed together with path plan or waypoints need to be transferred. The scenario manager will not interfere with own ship or the ANS algorithm after initial parameters are set.

Arrow 3 in Figure 3 indicates that the test evaluation module also needs to communicate with various control systems in own ship. All alarms, actions, in addition to information of ship positions, courses and speeds throughout the scenario are used by the test evaluation module for evaluating each test scenario. To do a full assessment of each test scenario, also course, speed and position for all target ships will need to be supplied to the test evaluation module, arrow 4 in Figure 3.

When performing simulation-based testing, the test interface needs to have the capacity to communicate all I/O between the control systems and the simulator at a rate that is sufficient for closed loop operation of the control system software. Normally, when using simulation-based testing in a Hardware-In-the-Loop (HIL) setup, it has been a requirement that the simulator must run in real time. In a HIL setup, the control system software is running on a Programmable Logic Controller (PLC) or similar with real time operating system. For more information on HIL, the reader is referred to (Johansen, Fossen, & Vik, 2005).

When testing ANS, it will be necessary to test large numbers of traffic scenarios of relatively long duration. If this should be tested in real time, the time consumption will be high. It is desirable to shorten the total test time as much as possible without sacrificing the test scope. This may be achieved either by using several test systems in parallel or have the simulator and the control system software under test run faster than real time, or a combination of these. For this to be possible, the control system software will have to run on emulated or virtual hardware, most probably in the cloud, where the simulation platform controls the simulation and computer clock cycle time.

Open Simulation Platform (OSP) (DNV GL, Kongsberg, SINTEF Ocean, NTNU, 2018) is a simulation platform which may potentially be used to ease the interfacing between the test management system and the virtual world with all its components. In addition, OPS will be running in the cloud, facilitating the possibility for control of the cycle time of the simulation and the virtual hardware. The platform is currently under development, and a short description is given in the following.

2.5 Open simulation Platform

OSP (DNV GL, Kongsberg, SINTEF Ocean, NTNU, 2018) is under development through a Joint Industry Project (JIP) with in total 24 participants, where Kongsberg, SINTEF Ocean, NTNU and DNV GL are the main partners.

The goal of the JIP is to develop a co-simulation platform to be used among ship designers, equipment and system manufacturers, yards, ship owners, operators, research institutes and academia. The co-simulation platform supports the functional mock-up interface (FMI) which is a tool independent standard to support both model exchange and co-simulation of dynamic models (FMI, 2019). Supporting the FMI standard will enable the users of the OSP to develop their simulation components in their known modelling environments. These components are then compiled to functional mock-up units (FMUs), before imported to the OSP for simulation-based testing. Each vendor may add their simulation components as FMUs making it easier to set up large simulations for a complete vessel with components and control software delivered by many different suppliers.

3 TEST SCENARIO EVALUATION

For evaluation purposes, COLREG, safety and other rules and regulations should be used. A lot of research has been conducted for path planning algorithms where COLREG is taken into use, and some examples are (Zhang, Yan, Chen, Sang, & Zhang, 2012), (Naeem, Irwin, & Yang, 2012) and (Campbell, Naeem, & Irwin, 2012). For evaluation of COLREG compliant ANS, not so much has been done. One of the most complete COLREG evaluation techniques has been developed by (Woerner, 2016), and both (Minne, 2017) and (Henriksen, 2018) are inspired by Woerner. (Stankiewicz & Mullins, 2019) have investigated both COLREG evaluation and adaptive scenario generation.

The COLREG are by purpose written such that seafarers need to use their judgement and common sense to interpret many of the rules. In order to practice good seamanship, also the autonomous vessels need to follow the COLREG, and vague rules may make it difficult to design the collision avoidance systems. The COLREG contain in total 38 rules divided into 5 parts in addition to four annexes. Not all parts of COLREG is possible to test using simulation-based testing and it is therefore important to clarify which of the COLREG rules that are covered by the ANS and included in the testing.

In the following, two different evaluation methods are described. Woerner (2016) has proposed a method where a total COLREG score, combined with a safety score and penalty scores for each part of the evaluation algorithms are calculated for each test scenario, and he has among others used court decisions for setting evaluation parameters. Another method is suggested by (Nakamura & Okada, 2019). By using relative distance between own ship and target ships and rate of change in bearing, the authors propose a method defining *Danger area*, *Caution area* and *Safety area* for bow and stern crossing and same way situation, which may be used in the evaluation of different test scenarios. The two methods are briefly described below in chapter 3.1 and chapter 3.2, respectively. More information may be found in the given references.

3.1 COLREG score and penalties

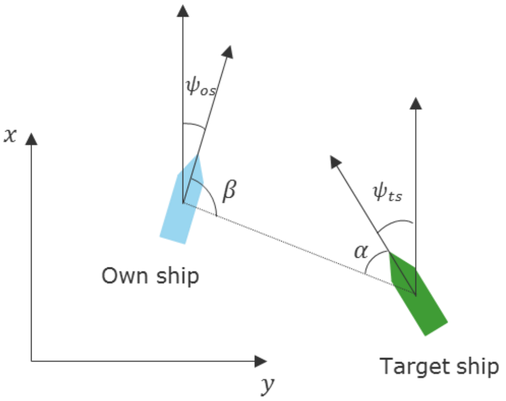


Figure 4: Pose between own ship and target ship

Pose between own ship and target ship may be used when evaluating different traffic situations, see Figure 4. Classifying the autonomous vessel which is targeted for testing, as *own ship* and other vessels as *target ships*, the pose between own ship and a target ship, is given by the relative bearing and contact angle. The contact angle α is the angle between the line of sight vector of target ship and the straight line between own ship and the target ship seen from the target ship. β is the angle between the line of sight of own ship and the straight line between target ship and own ship seen from own ship.

COLREG rule 14 is used as an example to describe the score and penalties method proposed by (Woerner, 2016). The rule shall prevent two vessels on nearly reciprocal courses from colliding, and the rule requires a port to port passing which may be evaluated using a combination of contact angle and relative bearing at closest point of contact (CPA). CPA is defined as the point on own ship's track where the range of the encounter between own ship and target ship is at its minimum. α_{cpa} and β_{cpa}, are defined as the contact and relative bearing between own ship and target ship

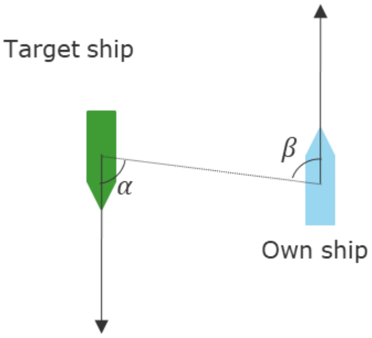


Figure 5: True port to port passing at CPA in head on situation

Figure 5 shows a true port to port passing at CPA, which is the preferred way of passing when in a head-on situation. A true port to port passing is achieved if $\alpha_{cpa} = -90°$ and $\beta_{cpa} = 270°$, and this needs to be reflected by the score function. Looking at α_{cpa}, one possible score function $S^{14}_{\alpha_{\mathrm{cpa}}}$ is

$$S^{14}_{\alpha_{\mathrm{cpa}}} = \left(\frac{\sin(\alpha_{cpa})-1}{2}\right)^2, \tag{1}$$

which may also be seen in Figure 6. The proposed score function gives maximum score at $\alpha_{cpa} = -90°$, while a starboard passing will result in 0 score.

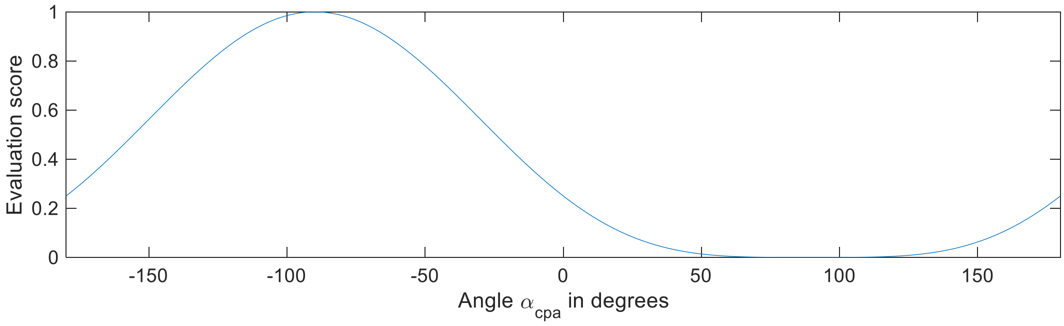


Figure 6: Plot of $S^{14}_{\alpha_{\mathrm{cpa}}}$

Similar score function may be used for β_{cpa}, and combining them gives the following score function for a true port to port passing:

$$S^{14}_{\Theta_{\mathrm{cpa}}} = S^{14}_{\alpha_{\mathrm{cpa}}} S^{14}_{\beta_{\mathrm{cpa}}} = \left(\frac{\sin(\alpha_{cpa})-1}{2}\right)^2 \left(\frac{\sin(\beta_{cpa})-1}{2}\right)^2. \tag{2}$$

The penalty score for evaluating the passing may then be given as

$$P^{14}_{\Theta_{\mathrm{cpa}}} = 1 - S^{14}_{\Theta_{\mathrm{cpa}}} = 1 - \left(\frac{\sin(\alpha_{cpa})-1}{2}\right)^2 \left(\frac{\sin(\beta_{cpa})-1}{2}\right)^2. \tag{3}$$

In a head-on situation, the rule requires a starboard manoeuvre to be commanded. (Woerner, 2016) did not propose a function for evaluating a non-starboard course change, therefore a new penalty function is proposed. By using the position of own ship at the time the target ship is detected, t_{0cpa}, as initial position, $\boldsymbol{p}_0$, and calculating a second position, $\boldsymbol{p}_2$ at t_2 assuming constant speed and heading, such that

$$\begin{aligned} t_2 &= 100\ t_{0cpa}, \\ \mathbf{a} &= \mathbf{p_0} - \boldsymbol{p}_2, \\ \mathbf{b} &= \mathbf{p} - \boldsymbol{p}_2, \end{aligned} \tag{4}$$

where $\boldsymbol{p}$ is the position of the own ship at any given time after the target ship has been detected, see Figure 7. If own ship for some reason is deviating from initial heading, the cross product between

$\boldsymbol{a}$ and $\boldsymbol{b}$ may be used to decide if own ship has deviated to port side or starboard side of the initial course. Using this together with

$$\mathrm{d} = \frac{\|\boldsymbol{a}\times\boldsymbol{b}\|}{\|\boldsymbol{a}\|}, \tag{5}$$

where d is the distance between the position of the own ship perpendicular to the line between the points $\boldsymbol{p}_0$ and $\boldsymbol{p}_2$, the penalty function P_{nsb}^{14} may be given as

$$P_{nsb}^{14} = \begin{cases} 1 & \mathrm{d} \geq d_{threshold} \; and \; \boldsymbol{a}\times\boldsymbol{b} > 0 \\ 1 - \left(\frac{2(d_{\text{threshold}} - \mathrm{d})}{d_{\text{threshold}}}\right)^4 & \frac{d_{threshold}}{2} < \mathrm{d} < d_{threshold} \; and \; \boldsymbol{a}\times\boldsymbol{b} > 0 \\ 0 & \boldsymbol{a}\times\boldsymbol{b} \leq \; 0 \text{ or } \mathrm{d} \leq \frac{d_{threshold}}{2} \end{cases}. \tag{6}$$

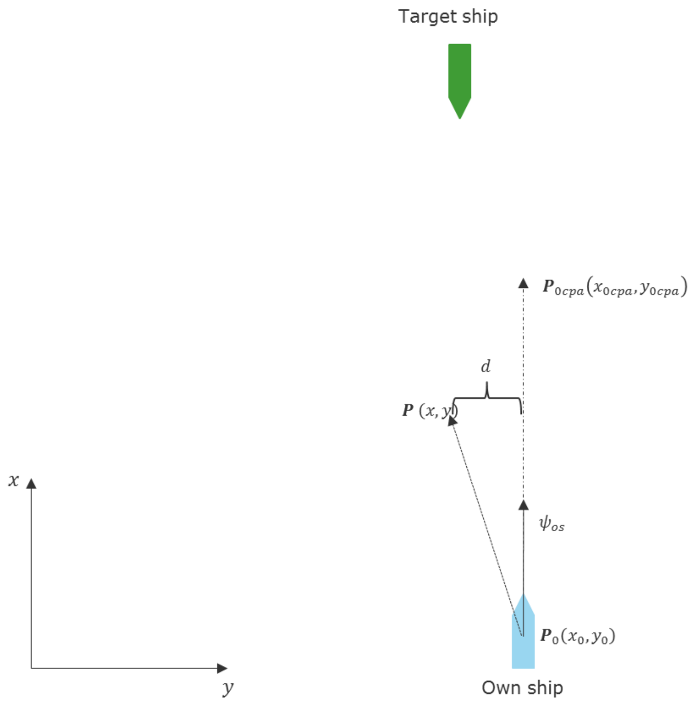


Figure 7: Head-on situation

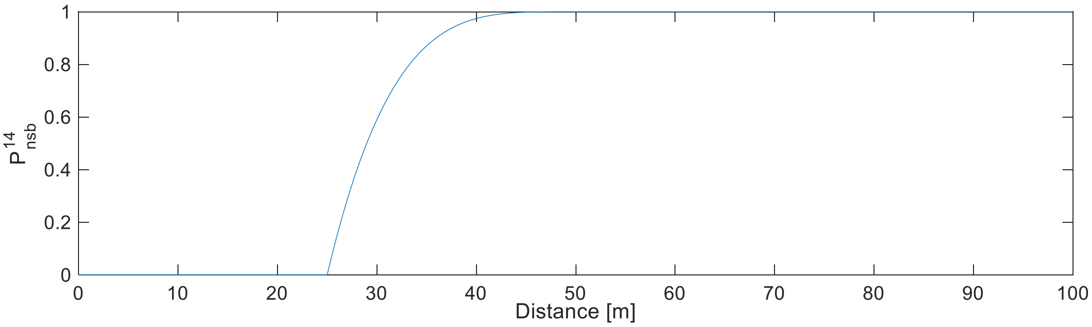


Figure 8: Penalty score for non-starboard course change

The penalty function is shown in Figure 8 for $\mathrm{d} \geq 0$ using $d_{threshold} = 50[m]$ for $\boldsymbol{a}\times\boldsymbol{b} > 0$. The total score for rule 14 is now given as

$$S^{14} = sat_0^1\left\{\left(1 - \gamma_{nsb}P_{nsb}^{14} - \gamma_{\Psi_{app}}P_{\Delta\Psi_{\text{app}}}^{8} - \gamma_{delay}P_{delay}^{8}\right)\left(1 - P_{\Theta_{cpa}}^{14}\right)\right\}, \tag{7}$$

where γ_{nsb}, $\gamma_{\Delta_{app}}$ and γ_{delay} are penalty coefficients which may be tuned.

3.2 Evaluation using anxiety estimation

Nakamura & Okada, (2019) proposes a method using anxiety estimation for evaluating an ANS towards COLREG. They have been collecting experience data where 12 captains and pilots were participating in navigational experiments. In total 135 encounters where simulated and 30 000 data points where collected.

According to the authors, the navigators use relative distance between ships, rate of change in bearing and crossing direction to recognize the risk of collisions with other ships. Due to these factors, they propose a set of evaluation diagrams, shown in Figure 9, where the diagram is divided into *Danger* area, *Caution* area and *Safety area* using relative distance and bearing change rate as input variables.

.

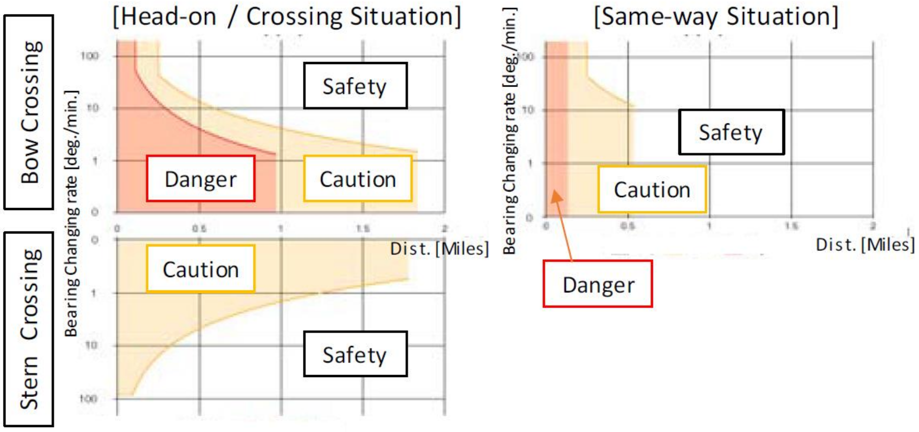


Figure 9: Evaluation area diagram (Nakamura & Okada, 2019)

The evaluation is done by summing the time used in the different phases in the evaluation area diagram. The time spent in the *Safety area* gives 0 penalty, while the time spent in *Caution area* and *Danger area* is multiplied with -1 and -2 respectively for penalty calculation. The authors propose to use the following equation to calculate the evaluation score for each scenario:

$$\text{score} = \frac{\sum_{t=0}^{t_{end}} - (2 \cdot Dangerous_t + 1 \cdot Cautionary)}{t_{end}} \tag{8}$$

The variable Dangerous is the period/time own ship was in the danger area during the scenario, while Cautionary is the period/time own ship was is in the caution area. t_{end} is the period/time of ship manoeuvring.

3.3 Final scenario assessment

A final assessment of the scenario evaluation needs to be taken by a human, but most probably it will not be feasible for a human operator to evaluate every one of the scenarios used for testing the ANS, especially not when testing is done in parallel and the test setup is running faster than real-time. The idea is that the test evaluation should trigger a human assessment. If, for example, a test result is below some threshold, the human operator should check the result and approve if acceptable. However, it is important to secure that the test evaluation algorithm does not let an actual ANS failure pass without a signalling the need for a manual check.

4 AUTOMATIC TEST SCENARIO GENERATION

One of the main challenges when it comes to implementing ANS, is to make the systems sufficiently safe, but what does that mean? An acceptance goal for autonomous system to be as safe or safer than conventional systems is challenging to prove, but one solution might be to test the algorithms in traffic scenarios that best represent the probable traffic scenarios a vessel might meet within e.g. 50, 100 or even 200 years of operation. In (Li, Huang, Liu, Zheng, & Wang, 2016), it is stated that existing testing approaches for autonomous vehicles in the automotive industry can be categorized into scenario-based testing and functionality-based testing. The authors argue that using either one of these methods is not enough, instead a combination of them should be used to design simulation-based tests for autonomous vehicles. The proposed method may also be used to design tests for autonomous ships, but in addition also robustness testing should be included. Robustness testing will demonstrate the ANS ability to handle errors, inaccuracies or noise in e.g. signals, sensors, actuators and equipment during operation. A possible test scope for the ANS could start with predefined, generic and stylistic single COLREG scenarios where own ship is only meeting one target vessel. For the next level of tests, the complexity of the scenarios could increase by introducing several target ships approaching own ship from different positions and with different heading such that own ship needs to handle several COLREG at the same time. The third level of tests could be location and operation specific test scenarios. Historical cases using AIS data and known incidents could be used as input to the generation test scenarios. For the last level of tests, automated scenario generation could be used. Automated, adaptive search for critical test scenarios is important to increase test coverage of the ANS. By using the evaluation of already performed test scenarios, adaptive search may be used to predict the most interesting (low score) test scenarios that might reveal a weakness in or failure of the tested system. The search algorithm can be based on some sort of sensitivity search through the test scenario scores, targeting potential weak spots using optimization or AI techniques like genetic algorithms, response surfaces, Bayesian optimization and Gaussian Processes (Machine Learning).

Immature systems would, using this approach, fail early while mature systems would fail late or eventually not fail, a strategy that can be considered agile and cost efficient. Re-testing of updated mature systems could be done using the same strategy. One could also envisage that even self-learning or adaptive ANS could be frequently or continuously tested in a similar way.

5 SIMULATION-BASED TESTING PROCESS

Simulation-based testing can be utilized in different phases of an ANS lifecycle process, such as during development, internal testing at manufacturer, or during formal testing. In this paper, the focus is on collecting evidence in a more formal assurance or certification process. Typically, formal processes involve different parties, such as system manufacturers, ship building companies, ship owners and verification organizations. The described simulation-based test system is specifically intended as part of the formal assurance process where the key parties are the ANS manufacturer, the end user of the ship and the verification organization, a role DNV GL or other class bodies could take.

Three critical aspects of the testing process are covered in the following, namely the cooperation between the ANS manufacturer and the verification organization, the aspect of independence and finally the validation of the test results.

5.1 Manufacturer cooperation

Manufacturer cooperation is key when performing testing of an ANS. The verification organisation depends on the manufacturer to be able to:

- secure correct software used for testing,
- interface their control system,
- understand how the ANS is working,
- commission the test setup and interface, and
- validate the test results

The last bullet in the above list is of high importance. As may be seen in chapter 3.3, the scenario evaluation should trigger a manual check of the test results by a human operator/tester in case of deviations. The human operator will then do an assessment of the scenario and flag this for follow up if necessary. All items flagged for follow up will then be discussed with involved parties, such as manufacturer, ship owner, class etc. If necessary, the scenario may be replayed while the manufacturer is checking their software.

Simulation-based testing does not require access to the source code of any part of the manufacturer ANS, since it is a black-box testing method, only considering the inputs and outputs of the SW-based parts. This can make it easier to cooperate with different manufacturers in a competitive business environment. Securing the IP of such manufacturer SW is also an important aspect of the OSP platform due to the same reasons.

5.2 Independence

Objectivity is important when testing software and the closer the developer is to the tester, the more difficult it is to be objective. The level of independence, and therefore the objectivity, increases with the ‘distance’ between the developer and the tester. The IEEE 1012 Standard for System and Software Verification and Validation (IEEE, 2012) defines three types of independence: *technical independence*, *managerial independence,* and *financial independence.* Technical independence means that the verification personnel or tools should not be involved or used in the development of the system. Managerial independence means that the verification organization should be independent from the system vendor organization, while financial independence means that the budget of the verification effort should be independent of the budget for the system development and delivery. The IEEE1012 also defines five forms of independence: *classical, modified, integrated, internal* and *embedded*, where classical independence Classical independence is when the verification organization is an external organization (different company), and embodies all three types of independence (technical, managerial, financial). This is the level of independence adequate when testing safety critical systems.

Manufacturers often have their own simulators which they use in development or internal testing of control system software. This also applies to the ANS manufacturers. It is possible to maintain classical independence even though the manufacturer simulator is used in the test setup. In such a setup, the verification organisation should provide a test interface between the control system subject to test and the simulator controlled by this control system. In this way, the verification organisation will have full control of all the signals interchanged between the simulator and the control system. In addition, the simulator should be validated by the verification organisation to be fit for purpose. Fit for purpose includes:

- simulators shall not set restrictions on the test scope and test scenarios,
- it shall be possible to get access to all relevant signals through the test interface,
- it shall be possible to validate the correctness of the simulator and all its components, and
- it shall be possible to validate the correctness of the interface between simulator, test interface and control system software.

5.3 Test result validation

It is crucial to validate the results from the use of the simulation-based testing to achieve the needed confidence in the test activity and finally in the correctness of the ANS under test. Apart from the fact that the ANS successfully should handle all the simulated scenarios according to the evaluation criteria, confidence arise from especially two aspects, being (i) correctness of the simulation-based test results and (ii) the sufficiency or completeness of the tested scenarios, i.e. the level of coverage.

Correctness of the simulation-based test results depend on the validation of the digital twin, such as the digital models, emulated systems, co-simulation of models and test interfaces. Validation is done in several ways and at different places in the testing process:

- interface and validation testing prior to starting the testing
- comparison of the digital twin simulation-based test results to results and data from real testing as mentioned in chapter 1
- cooperation with the manufacturer during testing where the manufacturer gives input whether the simulations and results are valid or trustworthy, as discussed in 5.1
- test results review activities performed by the manufacturer, ship owner and the verification organization, aiming at concluding and validating the end results of the testing activity

The final challenge of any testing is how sufficient, complete or representative the test scope is in addressing all the critical behaviour, functionality, robustness or performance of the system under test, i.e. the level of coverage. Confidence is often initially perceived by the absence of failed test results, but in the end, it is the level of coverage that finally creates the needed confidence. An ANS need to handle a very large number of different scenarios, and methods for assessing which scenarios that are representative or important to test and which are not, are a future research question relevant for many complex algorithms e.g. in autonomous or AI technologies.

6 CONCLUSION

For the autonomous ships to be accepted by the community, it is said that the autonomous ships need to be as safe or safer than conventional ships. Proving this may be a challenge, especially if only real-life testing is performed.

The autonomous navigation systems (ANS) should to go through a qualification scheme where safety goals and risks are identified, and qualification activities are performed to collect evidence for mitigating the risks and reaching the safety goals. DNV GL proposes to use a combination of real-life and simulation-based testing to assess the ANS. A Scenario Manager setting up test scenarios using a combination of scenario-based and functional based testing combined with robustness testing and automatic search for critical scenarios, will be a vital part of the Test System. Two different methods for evaluating the results from the testing are described. The Test Evaluation algorithm will need to trigger human assessment of possible ANS failures, and it is important that the evaluation algorithm does not fail to flag an actual ANS failure without signalling the need for a manual check.

Objectivity and independence are important factors when doing the final assessment of the ANS. Classical independence is when the verification organization is an external organization and embodies technical, managerial and financial independence. It is possible to maintain classical independence even though the manufacturer simulator is used in the test setup, as long as the test organization provides a test interface between the simulator and the control system and as long as the simulator is validated by the verification organisation to be fit for purpose.

REFERENCES

Campbell, S., Naeem, W., & Irwin, G. (2012). A review on improving the autonomy of unmanned surface vehicles through intelligent collision avoidance manoeuvres. *Annual Reviews in Control,*, (pp. 267–283).

DNV GL, Kongsberg, SINTEF Ocean, NTNU. (2018, September). *Open Simulation Platform*. Retrieved Augsut 29, 2019, from Open Simulation Platform: www.opensimulationplatform.com

DNVGL. (2017). *DNVGL-RP-A203: Technology Qualification.*

FMI. (2019). Functional Mock-up Interface. Retrieved June 2019, from https://fmi-standard.org

Heikkilä, E., Tuominen, R., Tiusanen, R., Montewka, J., & Kujala, P. (2017). Safety Qualification Process for an Autonomous Ship Protoype - a Goal-based Safety case Approach. *TransNav 2017 - 12th International Conference on Marine Navigation and Safety of Sea Transportation.*

Henriksen, E. (2018). *Automatic Testing of Maritime COllision Avoidance Methods with Sensor Fusion.* Master thesis, NTNU.

IEEE. (2012). *1012 Standard for System and Software Verification and Validation.*

Johansen, T., Fossen, T., & Vik, B. (2005). Hardware-in-the-loop testing of DP systems. *Dynamic Positioning Conference, Control Systems II.*

Li, L., Huang, W., Liu, Y., Zheng, H., & Wang, F. (2016). Intelligence Testing for AUtonomus Vehicles: A New Approach. *IEEE TRANSACTIONS ON INTELLIGENT VEHICLES*, (pp. 155-166).

Minne, P. (2017). *Automatic testing of maritime collision avoidance algorithms.* Master thesis, NTNU.

Naeem, ,. W., Irwin, G. W., & Yang, A. (2012). Colregs-based collision avoidance strategies for unmanned surface vehicles,”. *Mechatronics, vol. 22, no. 6,*, (pp. 669–678).

Nakamura, S., & Okada, N. (2019, March). Development of AUtomatic Collision Avoidance System and Quantitative Evaluation of the Maneuvering Results. *International Journal on Marine Navigation and Safety of Sea Trasportation, 13*(1), 133-141.

Rødseth, Ø. J., & Burmeister, H. C. (2015). Risk Assessment for an Unmanned Merchant Ship. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation 9(3)*, (pp. 357-364).

Safety4sea. (2018). *Human error the cause for most coastal vessels accidents in harbours.* https://safety4sea.com/human-error-the-cause-for-most-coastal-vessels-accidents-in-harbours/.

Stankiewicz, P., & Mullins, G. (2019). *Improving Evaluation Methodology for Autonomus Surface Vessel COLREGS Compliance.* MTS/IEEE OCEAN'19 Marseille Conference.

Vartdal, B.-J., Skjong, R., & St. Clair, A. L. (2018). *Remote-Controlled and Autonomous Ships.* Position Paper: DNV GL Group Technology & Research.

Woerner, K. L. (2016). *Multi-Contact Protocol-Constrained Collision Avoidance for Autonomous Marine Vehicles.* PhD thesis, Massachusetts Institute of Technology.

Wood, M., Robbel, P., Maass, M., Tebbens, R. D., Meijs, M., Harb, M., . . . Schlicht, P. (2019). *Safety First for Automated Driving.* Retrieved from https://newsroom.intel.com/wp-content/uploads/sites/11/2019/07/Intel-Safety-First-for-Automated-Driving.pdf

Zhang, J., Yan, X., Chen, X., Sang, L., & Zhang, D. (2012). A novel approach for assistance with anti-collision decision making based on the international regulations for preventing collisions at sea. (pp. 250-259). Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment.

International Seminar on Safety and Security of Autonomous Vessels
17 - 18 September 2019, Helsinki

Aalto University

Assessment of the Required Subdivision Index for Autonomous Ships based on Equivalent Safety

Jiri de Vos[1,*], **Robert Hekkenberg**[1]

[1] Department of Ship Design, Production & Operation, Delft University of Technology, the Netherlands

ABSTRACT

In recent years, a significant amount of research has been conducted on autonomous ships. Since it is assumed that these ships will sail with a significantly reduced crew or even without people on board, the design of the ship needs reconsideration. The absence of people on board and the associated safety measures could result in a more efficient design. However, to achieve the required design freedom, the existing regulatory framework will have to be amended. In this article, we will focus on potential changes in the Convention for Safety Of Life At Sea (SOLAS) and in particular on the Required Subdivision Index. The evaluation is performed by using the principle of equivalent safety, which will ensure that unmanned ships will be at least as safe as manned ships. The index gives a requirement for the allowed probability of sinking when a ship is damaged due to collision or contact. The safety level is related to the safety of ship, cargo, environment and crew. If the crew is no longer present, the consequences of an incident will be less severe, since the probability of casualties is no longer present. If the principle of equivalent safety is applied, a lower subdivision index can be accepted for unmanned autonomous vessels. In this article, the level of risk that a manned ship is subjected to will be derived by means of a risk analysis. In this risk analysis all logical consequences of a collision will be taken into account, covering both the probability of losing the entire ship and the consequences of the cases where the ship will not sink. Thereafter, the Required Subdivision Index for unmanned ships, which ensures an equivalent safety level to an equivalent manned ship, is established. The sensitivity of the result to changes in the data is discussed as well.

Keywords: Required Subdivision Index; SOLAS; Autonomous Ships; Risk Analysis; Equivalent Safety

1. INTRODUCTION

The research effort on autonomous ships has increased over the last years. The realisation of an autonomous ship will have as a consequence that the crew can be reduced significantly or even be removed entirely. Nevertheless, the business case of autonomous ships is still hard to make. As for most innovations within the maritime industry, the incentive for autonomous ships is economic efficiency (Karlis, 2018). Although there is a strong belief that autonomous ships would lead to more economic efficiency, only limited research has been performed in order to demonstrate what the overall effect of the change to autonomous shipping would have on transport costs (Frijters, 2017; Rødseth & Burmeister, 2015). More reductions in costs or improvement of transport performance for autonomous ships would make them more attractive and economically viable. Therefore, the design of the ship should be optimized for (unmanned) autonomous operations.

The design of a ship is subjected to regulations and requirements that limit the design freedom, but increase safety. Removing the crew from the ship reduces the risk of shipping, under the assumption that the probability that an incident occurs does not change, since the lives of the crew are no longer at risk. If the risk is lower, the requirements to the design of unmanned ships

---

[*] Corresponding author: phone, +31626737843, and email address, jiridevos@gmail.com

might become less strict, while maintaining equivalent safety. In this way more design freedom can be realised for unmanned ships and thus more economic efficiency.

The International Maritime Organization (IMO) is currently performing a regulatory scoping exercise (RSE) (IMO, 2018a). The objective has been defined as, “to assess the degree to which the existing regulatory framework under its purview may be affected in order to address Maritime Autonomous Surface Ship (MASS) operations”. This is an important step in the development for autonomous ships, since the result of the RSE will provide insight in “how safe, secure and environmentally sound” MASS operations need to be.

Other regulatory instances such as DNV GL and Bureau Veritas already shared their belief in the need for a new regulatory framework for autonomous ships. The development of a new regulatory framework would be the next step for IMO following the RSE. The regulatory instances have described what they believe the new regulatory framework should look like, but the proposals remain of a qualitative nature. There is only limited research being performed on defining the new regulations for autonomous ships.

The new regulations should ensure that autonomous ships will be as safe as manned ships. However, as stated before, this could lead to changes in the requirements that will create more design freedom for autonomous ships.

Within this article the required subdivision index will be evaluated and it is assessed how this index could be lowered, while still maintaining equivalent safety in case the ship is completely unmanned. In this article an approach is used to find the allowable reduction of the index for single ships.

In section 2 the method of the assessment is described. The basis of the method is derived from safety science, which will be elaborated upon first. The general approach is described as well. Next the concept of probabilistic damage stability and how this is used in the approach is explained. Thereafter, the determination of the consequences of damage is discussed. Last, the example ship that will be assessed is presented. In section 3 the results of the assessment are presented along with a discussion on these results. In section 4 the conclusions are presented. The recommendations follow in section 5.

2. METHOD

2.1. On equivalent safety

In order to be able to use equivalent safety for the assessment of the required subdivision index, the concept of safety must be understood. Safety is defined by the IMO as “Safety is the absence of unacceptable levels of risk (…)” (IMO, 2013). In other words, for something to be safe, it must be established what the acceptable level of risk is. Therefore, the assumption that the safety of autonomous ships should be equivalent to the safety of conventional ships means that both should be subject to the same level of risk. For this study, the damage stability-related level of risk of a conventional ship will be the benchmark for an unmanned autonomous ship of the same type and size.

Risk is defined as “a measure of the likelihood that an undesirable event will occur together with a measure of the resulting consequence within a specified time” (IMO, 2013). In other words, risk consists of two independent parts, a probability and a consequence. The probability is generally expressed as a probability per unit of time, for example per shipyear. The probability can be interpreted as “how often will the event happen (per unit of time)” or “how likely is it that the event will happen (per unit of time)”. The given number is usually between 0 and 1, meaning that an event will not happen and that an event will definitely happen respectively.

The consequences of the event can be of a different nature. For instance, the loss of human lives cannot directly be compared to the loss of a financial asset such as cargo. However, concepts such as the value of preventing a fatality (VPF) are used such that all consequences are expressed in monetary values. The following categories are taken as the possible consequences of a damaged ship:

- Loss of cargo
- Loss of fuel
- Damaged machinery
- Steel damage
- Loss of life
- Total ship loss

The loss of cargo, loss of fuel, damaged machinery and steel damage are considered for the damages where the ship remains afloat. If the damage leads to a total shiploss, these categories are incorporated in the consequences of a total shiploss. The determination of the values of the consequences is done is section 2.3.

Concluding, in order to find the damage stability-related level of risk, the following steps have to be taken. If it is known that the ship is damaged, the events that have to be evaluated are the damage cases that can occur. Each damage case has a probability of occurrence and a probability of survival. The determination of the damage cases and the probabilities is described in section 2.2. It can be determined which damage cases lead to each of the categories of consequences. For each category, the risk per damage case is determined by multiplying the probability of occurrence with the consequences of that category. The total risk per category is the summation of the risk of that category per damage case. The overall damage stability-related level of risk is the summation of the risk per category.

For the transition towards an unmanned autonomous ship, the overall level of risk is reduced with the risk of loss of life, when it is assumed that the design remains unchanged. Since this lowers the overall level of risk, changes to the unmanned autonomous ship can be allowed. The changes should result in a change of the probability of occurrence for the remaining categories of consequences. This is further described at the end of section 2.2. The costs of the consequences are assumed to remain unchanged.

2.2. Probabilistic damage stability

The requirement concerning damage stability is called the required subdivision index (referred to as index R). The attained subdivision index of a ship (referred to as index A) has to be higher than index R. The definitions of index R and A are described in SOLAS (IMO, 1980).

The index A is a property of the ship and can be considered as a total probability of survival, given that the ship is part of a collision (Papanikolaou & Eliopoulou, 2008). Thus it reflects the ships capability to survive a collision or contact that leads to damage to the hull. The index A is calculated by evaluating most of the possible damage cases that follow from collision or contact.

A damage case is a situation where one or more adjacent compartments are flooded. The length of the damage of a certain damage case corresponds to the overall length of the compartments under consideration. The height of the damage corresponds to the height of the bulkhead deck. The depth of the damage corresponds to the minimum depth of the compartments under consideration. The probability of occurrence of the damage cases is derived from a study by Lützen on ship collisions (Lützen, 2001). SOLAS prescribes a method to calculate the probability of occurrence for the specific damage case (p_i).

The flooding of the compartments has an influence on the stability of the ship. The new stability properties are used to calculate a probability of survival for the specific damage case (s_i). Together with the probability of occurrence, this number is used to calculate index A.

$$A = \sum_i p_i * s_i$$

The ship is considered in three loading conditions. The deepest subdivision draught (d_s) is the waterline which corresponds to the Summer Load Line draught of the ship. The light service draught (d_l) is the service draught corresponding to the lightest anticipated loading and associated tankage, including such ballast as may be necessary for stability and/or immersion. The partial subdivision draught (d_p) is the light service draught plus 60% of the difference between the light service draught and the deepest subdivision draught. The total index A consist of three partial indices (A_s, A_p and A_l) corresponding with the three loading conditions as follows:

$$A = 0.4A_s + 0.4A_p + 0.2A_l$$

Subsequently, the index A has to be higher than the prescribed index R. If the length of the ship (L_S) is over 100 meters, the index R is defined as:

$$R = R_0 = 1 - \frac{128}{L_S + 152}$$

If the length of the ship is less than 100 meter but greater than 80 meter, the index R is defined as:

$$R = 1 - \frac{1}{1 + \frac{L_S}{100} * \frac{R_0}{1 - R_0}}$$

If a ship is shorter than 80 meter, there is no requirement concerning its subdivision index.

The method of finding the probability of occurrence and the probability of survival for the damage cases is used in the risk analysis as described in section 2.1. A lower index R for a ship of a certain type and size gives the possibility to reduce the index A. If the index A changes, the probability of occurrence and the probability of survival of the damage cases also change. Subsequently the overall level of risk of the ship also changes.

Within the approach that is described in this article, it will be assumed that all probabilities will change with the same rate. The rate is defined as $\frac{A_u}{A_m}$, where A_m is the index A of the manned ship under consideration and A_u is the index of the unmanned autonomous ship, of the same type and size, that results in the same level of risk. By using a solver the value of A_u can be found. The differences between A_m and A_u is the allowable change in the index R for the considered ship of a certain type and size.

Small reductions of the index A can be realised by reducing the minimum GM the ship is allowed to sail with or by reducing the number of tanks in the ship. These changes can already lead to more transport efficiency. Even more transport efficiency can be realised if larger reductions of the index A are allowed.

2.3. Determination of consequences

As was mentioned before, the consequences for a damaged ship depend on the damage case that occurs. For any damage case, if the ship remains afloat, the consequences are a combination of one or more of the following categories: loss of cargo, loss of fuel, damaged machinery and steel damage. If the ship sinks, these consequences will occur as well and they are incorporated in the costs of a total ship loss. The loss of life is evaluated separately.

2.3.1 Loss of cargo

The loss of cargo will occur when a cargo hold is penetrated and the ship remains afloat. The loss of cargo when the ship is lost is incorporated in the consequences of a total shiploss. The risk of losing cargo is calculated by establishing the damage cases that lead to the penetration of a cargo hold, while the ship remains afloat. The risk per damage case is the probability that the damage case occurs multiplied with the costs of the loss of cargo. The total risk of losing cargo is a summation of the risk of all the relevant damage cases.

The worst case scenario is evaluated, where it is assumed that all cargo in and above a penetrated cargo hold is considered to be lost. Different types of cargos lead to different cargo values. E.g. containers are much more valuable than dry bulk. The most transported dry bulk by ship are coal, iron ore and grain, accounting for nearly two thirds of the dry bulk trade (Chen, 2017). Of these three commodities the most valuable is grain. Its current value is €185 per tonne, which is about three times higher than the value of coal and iron ore (“Wheat vs Coal,” 2019; “Wheat vs Iron Ore,” 2019). The average value (€40,000 (IHS Markit, 2017)) and maximum weight (24 tonnes) of a TEU would lead to a minimum value of around €1,600 per tonne.

For the purpose of this risk analysis, it will conservatively be assumed that the ship will transport containers. The maximum number of containers a ship can transport will be used as the amount of cargo on board. The value per TEU will be taken as €40,000 (IHS Markit, 2017). In partial loading conditions, 60% of the capacity of each cargo hold is used.

2.3.2 Loss of fuel

If a fuel tank is penetrated, the fuel will flow out and that would be a threat to the environment. The fuel would need to be cleaned up, which will include costs. The risk of losing fuel is calculated by establishing the damage cases that lead to the penetration of a fuel tank, while the ship remains afloat. The risk per damage case is the probability that the damage case occurs multiplied by the costs of the loss of fuel. The total risk of losing fuel is a summation of the risk of all the relevant damage cases.

The costs of losing fuel are estimated using the size of the spill by $€37,819 * V^{0.7233}$ (IMO, 2018b). The value of the fuel that is lost is much lower than the clean-up costs and is incorporated in the uncertainty of the actual value of the clean-up costs. As will be discussed in section 3, the sensitivity of the result to the loss of fuel is low. Therefore a more accurate estimation is not needed. If the damage case will cause the ship to sink, the clean-up costs are incorporated in the costs of a total ship loss.

2.3.3 Damaged machinery

When the engine room is penetrated, while the ship remains afloat, the machinery will be damaged. The risk of damaged machinery is calculated by establishing the damage cases that lead to the penetration of the engine room, while the ship remains afloat. The risk per damage case is the probability that the damage case occurs multiplied by the costs of damaged machinery. The total risk of damaged machinery is a summation of the risk of all the relevant damage cases.

The cost estimation of the damaged machinery is based on the costs of a new drive train. Aalbers provides a cost estimation for the entire drive train of $€4{,}200 * P^{0.79}$, with P the installed power (Aalbers, n.d.). As will be discussed in section 3, the sensitivity of the result to damaged machinery is low. Therefore a more accurate estimation of the costs of damaged machinery is not needed and spills of polluting liquids such as lube oil or black water are not incorporated.

2.3.4 Steel damage

After a collision where the ship remains afloat, the damages to the ship will have to be repaired before the ship can be used again. Each damage case where the ship remains afloat will have steel damage as a consequence. Per damage case the risk of steel damage is calculated by multiplying the probability of the damage case with the relevant costs of the repairs. The total risk of steel damage is a summation of the risk of all the relevant damage cases.

In order to perform the repairs the ship would need to go into a dry-dock. Aalbers (Aalbers, n.d.) provides an estimation of the costs of dry-docking of 1-2% of the newbuilding price of the ship, while Hansen (Hansen, 2013) shows that the actual costs of dry-docking are often underestimated. Therefore, conservatively, the costs of dry-docking are estimated as 3% of the newbuilding price.

Next to the costs of dry-docking, the costs of repairs are estimated per meter of damage. The amount of steel per meter of ship length is estimated by dividing the ship's steel weight by the ship length. The actual amount of steel that needs to be replaced depends on the penetration depth of the damage. If only the outer hull is damaged, it is assumed that this corresponds to 1/8 of the cross-section. If the inner hull is damaged too, it is assumed that this corresponds to 1/4 of the cross-section. By using material costs of €850 per tonne of steel (Aalbers, n.d.) and an estimation of 300 required man-hours per tonne of steel (Butler, 2013), the costs of the repairs per meter of damage are calculated as follows:

$$Cost_{repairs} = €14{,}500 * \frac{steel\ weight}{ship\ length} * \left(\frac{1}{8}\ or\ \frac{1}{4}\right)$$

The total costs of steel damage per damage case is the costs of the dry-dock plus the costs of the repairs of the damage.

2.3.5 Loss of life

Crew members that are present on a ship that is part of a collision are subjected to the potential of losing life. The loss of life can be compared with other risks by using the VPF. The VPF is a value that represents society's willingness to pay for small reductions of the probability of losing life. According to EMSA, the VPF is approximately €6.25 million per fatality (European Maritime Safety Agency, 2015b). The risk of losing life is calculated by multiplying the probability of losing life with the VPF.

In order to find the probability of losing life during a collision or contact, data on ship accidents from 2000 to 2012 is used (Eleftheria, Apostolos, & Markos, 2016). The data by Eleftheria et al. is a collection and overview of the data available on collisions and fatalities. From this data the statistical average loss of life per accident (SALL) can be derived for general cargo ships, bulk carriers and containerships. The SALL is determined by dividing the number of fatalities by the number of accidents (see Table 1).

As can be seen in Table 1, the SALL differs per ship type. This might be explained by the different average size of each ship type. Bulk carriers and containerships are generally much larger than general cargo ships (Equasis, 2012), thus providing a safer environment for the crew in case of a collision. As will be described in section 2.4, the effect of removing crew on the total level of risk is expected to be largest for smaller ships. Therefore, the accident data of general cargo ships is used.

Table 1: Finding the statistical average loss of life during collision or contact for general cargo ships, bulk carriers and containerships.

		General Cargo	Bulk carrier	Containership
Fleet at risk		118,325	67,822	45,099
Collision or contact	Per shipyear	7.471E-03	7.472E-03	9.383E-03
	Total	884	507	423
Fatalities during collision or contact	Per shipyear	1.881E-03	1.920E-04	8.870E-05
	Total	223	13	4
Statistical average loss of life		0.252	0.026	0.009

In the data by Eleftheria et al. (2016) there is no distinction between fatalities when the ship was lost or stayed afloat. The lack of data on this subject makes it impossible to determine the cause of the fatalities during collision or contact at this point. The SALL in Table 1 has been calculated with the assumption that fatalities occur evenly over all accidents. However, if the fatalities would only occur when the ship is lost this would have an impact on the analysis. The other extreme is when the fatalities only occur when the ship is not lost. In Table 2 the SALL for the three interpretations of the data is presented for general cargo ships. The impact of these interpretations on the result will be evaluated in section 3.

Table 2: The SALL for general cargo ships when the data is interpreted in three different ways.

	Fatalities occur evenly	Fatalities occur when ship is lost	Fatalities occur when ship is not lost
Fatalities	223	223	223
Ship accidents considered	884	82	802
Statistical average loss of life	0.252	2.720	0.278
Probability of occurrence of accidents	1	1 – A	A

Concluding, the risk of losing life is calculated by multiplying the SALL with the VPF. The VPF is taken as €6.25 million and the SALL as 0.252, corresponding to the accident data of general cargo ships where the fatalities occur evenly over all accidents.

2.3.6 Total ship loss

The risk associated with a total ship loss is calculated by multiplying the probability of a total ship loss (1 minus index A) with the costs of a total ship loss. The costs resemble the possible consequences if the ship remains afloat, but are represented by loss of cargo, loss of ship and wreck removal costs (including clean-up of any fuel spill). The costs related to the potential loss of life are incorporated in the category “loss of life”.

The value of the cargo on board of the ship will be lost and the calculations are the same as in section 2.3.1. Also, evidently, the ship is lost and the ship has a certain value as well. It is assumed that ships are depreciated over their entire lifetime towards their scrap value of a minimum of €190 per LDT (Jain, 2017). Since this is a study on the potential of losing the ship, it is assumed that on average ships are lost halfway their expected lifetime. Therefore, the value of the ship is taken as halfway its depreciation.

The wreck will have to be removed and cleaning of the environment will be necessary in order to prevent damage to the environment. The costs related to these activities are highly dependent on the circumstances of the accident. However, EMSA provides an estimate of one to three times the newbuilding price of the ship (European Maritime Safety Agency, 2015a). In this research, two times the newbuilding price will be taken as costs for wreck removal.

2.4. The ship

It is expected that the changes in the requirements concerning damage stability are largest for smaller ships. When the ship becomes larger, the size of the crew does increases with a lower rate compared to the amount of cargo, installed power or capital costs. Therefore, it is expected that the contribution of the crew to the overall level of risk is lower for larger ships than for smaller ships.

The method that is described in the previous paragraphs will be used to assess a 4,000 ton deadweight general cargo ship. All the particulars that are needed to determine the consequences of any damage case are presented in Table 3. The ship has one cargo hold. The engine room is located in the aft part of the ship. The ship has three fuel tanks, of which one is located next to the engine room on portside. The other two are located in the double hull in the middle of the ship.

Table 3: The particulars of the ship that is evaluated in this article.

Ship type	General cargo
Length	89.9 m
Lightweight	1503 t
Steel weight	1020 t
DWT	4050 t
TEU	218
Crew	10
Installed power	1500 kW
Fuel oil	308 t
Newbuilding price	€7 million
Required subdivision index	0.444
Attained subdivision index	0.445

3. RESULTS AND DISCUSSION

The assessment of the ship described in section 2.4 leads to the risk profile of the ship as presented in Table 4. From this overview it can be seen that risk of a total ship loss is the main contributor to the damage stability-related overall level of risk. The risk of losing life also has a significant contribution. The remaining four categories, however, have a contribution of 1% or less. Thus even if these categories are underestimated with a factor two, the risk profile of the ship changes only little. The risk of losing cargo is even zero. The reason is that this ship has only one cargo hold. If the cargo hold is penetrated, the probability of survival is always zero. The contribution of the loss of cargo related to a total shiploss, however, is significant and will increase with the size of the ship. A more accurate estimation of the costs of loss of fuel, damaged machinery and steel damage is not required.

Table 4: Overview of the risk profile of the ship under evaluation in its conventional form as a manned ship.

Type	Risk	Probability	Contribution to the overall level of risk
Loss of cargo	€ -	0	0.0%
Loss of fuel	€ 174,000	0.161	1.1%
Damaged machinery	€ 56,000	0.041	0.4%
Steel damage	€ 206,000	0.445	1.3%
Loss of life	€ 1,577,000	0.252	10.2%
Total ship loss	€ 13,479,000	0.555	87.0%
Overall level of risk	€ 15,491,000		
Attained subdivision index		0.445	

Using the approach described in this article, the risk profile of an unmanned autonomous ship of the same type and size is found. The results are presented in Table 5. As can be seen, the risk of total ship loss increases, since the probability on losing the ship increases when index A is reduced. The overall level of risk is mainly determined by the risk of a total ship loss. The unmanned autonomous ship should have an index A of 0.378 to be subjected to the same level of risk as the manned ship. This is a reduction of 0.067 or 15.2%.

Therefore, if the index R for the unmanned autonomous ship would be 0.378, it will be ensured that it will have equivalent safety compared to the manned ship.

Table 5: Overview of the risk profile of the ship under evaluation in its revised form as an unmanned ship.

Type	Risk	Probability	Contribution to the overall level of risk
Loss of cargo	€ -	0	0.0%
Loss of fuel	€ 148,000	0.137	1.0%
Damaged machinery	€ 47,000	0.035	0.3%
Steel damage	€ 175,000	0.378	1.1%
Loss of life	€ -	-	-
Total ship loss	€ 15,122,000	0.622	97.6%
Overall level of risk	€ 15,491,000		
Attained subdivision index		0.378	

As described in section 2.3.5, uncertainties are present in the accident data and thus the risk of losing life. In Table 6 the resulting new index A of the unmanned autonomous ship is presented if the approach described in this article is used with different values for the risk of losing life. The results in Table 5 correspond to the results in the column 'general cargo ship – fatalities occur evenly' of Table 6.

The results in Table 6 show that the allowable change in the index varies significantly, depending on the cause of the fatalities. The results also show that the differences per ship type have a significant effect on the outcome. Therefore, further research to reduce the uncertainties is needed and are described in section 5.

Table 6: The allowable changes in required subdivision index for different interpretations of the accident data. The results under general cargo ship use different assumptions for the cause of fatalities. The result under containership assumes that fatalities occur evenly over all accidents.

	General cargo ship			Containership
	Fatalities occur evenly	Fatalities occur when ship is lost	Fatalities occur when ship is not lost	Fatalities occur evenly
SALL	0.252	2.720	0.278	0.009
Risk of losing life	€ 1,577,000	€ 9,428,000	€ 774,000	€59,000
A_{new}	0.378	0.041	0.412	0.443
Change	-0.067	-0.404	-0.033	-0.002
%	-15.2%	-90.8%	-7.5%	-0.6%

4. CONCLUSIONS

The assessment of the 4,000 ton deadweight ship shows that the risks associated with a total ship loss and loss of life are the main contributors to the damage stability-related level of risk. Therefore, removing the crew reduces the overall level of risk significantly for autonomous ships.

Subsequently, based on equivalent safety, the required subdivision index can be lowered for unmanned autonomous ships. However, as can be seen in the results, the size of the reduction depends strongly on missing accident statistics concerning the loss of life. Further research to reduce these uncertainties is described in the recommendations.

Even small reductions of the required subdivision index might already lead to an increase in transport capacity by reducing the minimum GM the ship is allowed to sail with. For larger reductions in the required subdivision index this effect can be extended by a simpler and more efficient design.

5. RECOMMENDATIONS

There seems to be a discrepancy between the theoretical probability of survival and the probability of survival that can be derived from accident data. The theoretical probability of survival of a ship is equal to the attained subdivision index, which is lower than 0.7 for ships under 275 metres and thus for most ships. Therefore, it is expected that at least 30% of the accidents concerning collision or contact should lead to a total ship loss. From accident data it can be derived that only 10% or less of the accidents concerning sea going cargo ships lead to a total ship loss, depending on the type of ship. It should be further investigated why the theory differs from the reality. Therefore it is recommended to perform a study on cases of collision and contact. Within this study it should be derived what the theoretical probability of survival was after the ship was damaged. This should indicate whether all ships that should have been lost in theory actually were lost and whether all ships that should have survived in theory actually survived.

The accident data that is available suggests that the potential loss of life depends on the type of ship. The loss of lives is significantly lower for bulk carriers and container ships than for general cargo ships. This could be the cause of the average size of the ships in each category. General cargo ships are generally smaller than bulk carriers and container ships. Further investigation on the influence of the size of the ship on the potential loss of life is needed. It is, therefore, recommended to collect data on the size of the ships in the accident data and on what size of ship a fatality occurred.

Furthermore, the relation between the size of the crew and the risk of losing life is unknown. It is recommended to investigate if the casualties occurred incidentally over all accidents, regardless of the size of the crew, or if the risk of losing life is associated with the risk of losing the entire crew.

This research focusses on the events and consequences that assume that a ship is damaged as a result of collision or contact. The probability that a ship is part of a collision or contact is not taken into account. It may well be that the probability that a ship is part of a collision will change if the transition towards unmanned ships is made. If this probability increases, a higher survivability of the ship might be required. If this probability decreases, an even lower survivability might be required. It is recommended to further investigate how the probability that a ship is part of a collision will change for unmanned ships.

ACKNOWLEDGEMENTS

The research leading to these results has been conducted within the NOVIMAR project (NOVel Iwt and MARitime transport concepts) and received funding from the European Union Horizon 2020 Pro-gram under grant agreement n° 723009. The design and data of the ship used in this research has been provided by DEKC Maritime. The authors would like to thank Herbert Koelman, Lector Maritime Innovative Technologies at Maritiem Instituut Willem Barentsz and director of SARC, for his support in the understanding of and working with probabilistic damage stability.

REFERENCES

Aalbers, A. (n.d.). Evaluation of ship design options.

Butler, D. (2013). *A Guide to Ship Repair Estimates in Man-hours*. Elsevier. https://doi.org/10.1016/C2011-0-07776-1

Chen, J. (2017). Dry Bulk Commodity. Retrieved May 6, 2019, from https://www.investopedia.com/terms/d/dry-bulk-commodity.asp

Eleftheria, E., Apostolos, P., & Markos, V. (2016). Statistical analysis of ship accidents and review of safety level. *Safety Science*, *85*, 282–292. https://doi.org/10.1016/j.ssci.2016.02.001

Equasis. (2012). The World Merchant Fleet in 2012.

European Maritime Safety Agency. (2015a). *Impact assessment compilation part 1; Impact assessment in accordance with the EC IA*.

European Maritime Safety Agency. (2015b). *Risk Acceptance Criteria and Risk Based Damage Stability. Final Report, part 1: Risk Acceptance Criteria*.

Frijters, T. (2017). *Future Ships*. TU Delft.

Hansen, K. F. S. (2013). *Analysis of estimations, quotations and actual costs related to dry-docking*.

IHS Markit. (2017). *Vessel Accumulation and Cargo Value Estimation*.

IMO. (1980). *International Convention for the Safety of Life at Sea (SOLAS), 1974 as amended*. IMO.

IMO. (2013). *Guidelines for the Approval of Alternatives and Equivalents as Provided for in Various IMO Instruments*. *MSC.1/Circ.1455*. IMO.

IMO. (2018a). *Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS)*.

IMO. (2018b). *Revised Guidelines for Formal Safety Assessment (FSA) for Use in the IMO Rule-Making Process*. *MSC-MEPC.2/Circ.12/Rev.2*. IMO.

Jain, K. P. (2017). *Improving the Competitiveness of Green Ship Recycling*. Delft University of Technology.

Karlis, T. (2018). Maritime law issues related to the operation of unmanned autonomous cargo ships. *WMU Journal of Maritime Affairs*, *17*(1), 119–128. https://doi.org/10.1007/s13437-018-0135-6

Lützen, M. (2001). *Ship Collision Damage*. *Technical University of Denmark*. https://doi.org/10.1111/j.1365-2958.2006.05502.x

Papanikolaou, A., & Eliopoulou, E. (2008). On the development of the new harmonised damage stability regulations for dry cargo and passenger ships. *Reliability Engineering and System Safety*, *93*(9), 1305–1316. https://doi.org/10.1016/j.ress.2007.07.009

Rødseth, Ø., & Burmeister, H.-C. (2015). D10.2: New Ship Designs for Autonomous Vessels, (First outline).

Wheat vs Coal, South African export price - Price Rate of Change Comparison. (2019). Retrieved May 6, 2019, from https://www.indexmundi.com/commodities/?commodity=wheat&commodity=coal-south-african

Wheat vs Iron Ore - Price Rate of Change Comparison. (2019). Retrieved May 6, 2019, from https://www.indexmundi.com/commodities/?commodity=wheat&commodity=iron-ore

Empirical analysis of complex network for marine traffic situation

Zhongyi Sui [1,*], Yuanqiao Wen [2,3], Chunhui Zhou [1,2,4], Changshi Xiao [1,2,4], Fan Zhang [1,2,4], Liang Huang [2,3] and Hualong Chen [1]
[1] (School of Navigation, Wuhan University of Technology, China)
[2] (Intelligent Transportation Systems Research Center, China)
[3] (National Engineering Research Center for Water Transport Safety, China)
[4] (Hubei Key Laboratory of Inland Shipping Technology, China)

ABSTRACT

With the increasing water transportation, the maritime management departments need to improve traffic service to meet the development needs of water transportation. Therefore, the quantitative expression of marine traffic situation has attracted attention of scholars. At present, there is no macro index to indicate the complexity of marine traffic situation. The work studied the complexity of marine traffic situation based on the theories of complex network and network dynamics. The ship was regarded as the vertex, and the relationship between the ships as the edge. Moreover, the topological characteristics such as degree, vertex strength, clustering coefficient and network structure entropy were used to statistically characterize the evolution of marine traffic situation. The actual scenario analysis was used to reveal the change of marine traffic situation. The results showed that the complex network can provide an intuitionistic and accessible metric to reflect the marine traffic situation.

Keywords: Marine traffic management; Marine traffic situation; Complex network; Topological characteristic.

1. INTRODUCTION

Marine traffic system is composed of static, dynamic, informative and organizational elements such as human, ship and navigation environment. The marine traffic situation consists of the macro and micro traffic behaviors and their evolution of the marine traffic objects (ships) in the system under the constraints of traffic environment and traffic management rules, reflecting the operation state and development trend of marine traffic system. The quantitative expression of situation provides an effective method for intelligent analysis to understand the complexity and operation characteristics of marine traffic system. In December 2004, the International Maritime Organization(IMO) required all vessels over 299GT to install an automatic identification system(AIS) transponder on board. Besides, the automatic identification system is mostly used in ship navigation, collision avoidance and so on [1-5]. With the increasing number of ships equipped with AIS, it has become a focus issue to make full use of ship AIS data for the research of marine traffic situation.

Recently, the navigation environment is becoming more and more complex, with the increasing risk of marine traffic. This also increases the requirements and challenges for traffic controllers and ship operators. Under a low complexity situation, the ship operators have enough time to resolve conflicts, and can minimize the disturbance under the premise of navigation safety. On the contrary, under a high complexity situation, ship operators will ignore some potential conflicts. Therefore, it is urgent to understand the marine traffic situation. When two or more ships are approaching, controllers and ship operators can take the corresponding solutions immediately. Hence, it is important to descript the between-ship relation and the difficulties brought by marine traffic situations to controllers and seaman [6].

The characterization of marine traffic situation based on traffic flow complexity is the core of situation awareness (SA). More ships and closer distance between them cause greater pressure

---

[*] Corresponding author: Mobile:+8618696118075 Email:122643563@qq.com

for ship operators. So the traffic density is the most intuitionistic indicator reflecting marine traffic complexity. Besides the traffic density, the other traffic flow characteristics include ship speed, course, and the proportion of ship trajectory [7-9].

With the installation of equipments (Radar, Automatic Identification System(AIS), etc.), it is easier to obtain the traffic flow information. However, the massive data of traffic flow are not enough to help controllers or seaman understand the current marine traffic situation. In some cases, data redundancy leads to the wrong decisions made by maritime controllers or seamen. Thus, some researchers focus on finding the relationship between the traffic flow characteristics and marine traffic situation by traffic statistics. The average speed, quantity of traffic flow, traffic distribution, types of ship, etc. are treated as the basic features of the traffic system [10-12].

Some researchers focus on the relationship between statistical parameters of traffic flow and traffic situation, showing the long-term traffic situation [13, 14]. Some other scholars build the simulation models based on historical data to study traffic situation [15-17]. Recently, some researchers use the machine learning for situational awareness [18-23]. They do not pay attention to the real-time marine traffic situation. Therefore, the complexity of marine traffic situation should be studied from the perspective of complex systems. With the developed mathematical description of marine traffic situation, the factors of traffic density and traffic conflict are used to describe the relationship between two ships [6]. However, in the complex traffic systems, along with the increase ships and the decreasing ship distance, the complexity of multi-ship is increasing with non-linear relationship. It is unreasonable to superimpose the complexities of all ships

The above research studies the complexity of marine traffic situation from different perspectives, but ignores the between-ship proximity from the perspective of structure. In fact, the structure of marine traffic provides a lot of information among two or more ships. At different times, the marine traffic system shows different structures.

In order to research the structure of marine traffic situations, the complex network theory were introduced to describe the marine traffic situation, establishing the weighted network models corresponding to marine traffic situations at different points in time. From a systematic perspective, the structural properties of marine traffic situations were investigated based on AIS data.

The contributions of the work are as follows. First, the proposed concept of marine traffic situation structure provides a new approach to describe the marine traffic situation from the cognitive perspective. Secondly, the marine traffic is extracted into a complex network to analyze the topological characteristics based on the complex network of marine traffic situation. The dynamic changes of complex network reflects the complexity of marine traffic situation and its evolution process.

The rest of the work is organized as follows. In Section 2, we introduce the complex network theory, and weighted the complex network model of marine traffic situation. Section 3 calibrates the analysis of topological characteristics in the weighted complex network of marine traffic situation. Moreover, we analyze the marine traffic network from vertex strength, connection rate, network density and network structure entropy. In section 4, the conclusions are addressed.

2. COMPLEX NETWORK OF MARINE TRAFFIC SITUATION

2.1. Complex network theory

Complex networks can describe the complex problems in natural science, social science, management science and engineering technology. It takes the mathematics, statistical physics, computer science and other sciences as analysis tools, and the complex systems as research objectives. Complex network is a method for extraction and description of complex systems, which highlights the topological characteristics of system structures [24-26]. In real life, complex systems can be described by the complex networks, such as communication networks, social networks and transport networks. Generally, any complex system containing abundant component units can be regarded as a complex network when the component units are interrelated.

Complex network is a research object based on graph theories. A graph G is the ordered pair $\left(V(G),E(G)\right)$ consisting of a node set $V(G)$ and edge set $E(G)$, disjoint from $V(G)$, of edges, together with an incidence function ψ_G that associates with each edge of G an unordered pair of (not necessarily distinct) vertices of G. If e is an edge, u and v are vertices such that $\psi_G(e)=\{u,v\}$, then e joins u and v, and the vertices u and v are the ends of e. We denote the numbers of vertices and edges in G by $v(G)$ and $e(G)$, which are called the order and size of G, respectively.

Essentially, a marine traffic system is a complex system from the perspective of complex networks. With ship regarded as node, the work expressed the relationship between ships as undirected edge, and then established a corresponding complex network describing marine traffic situation. With the movement of ships, the complex network of marine traffic situation constantly changed. Therefore, the dynamic change of network parameters reflects the change of marine traffic situation in an area.

2.2. Connection method in marine traffic situation network

As a complex system, the structure of marine traffic, as well as the complex relationship within the system, has always been the focus. In many cases, controllers need to perceive the marine traffic situation in an area. Therefore, it is necessary to express the macro situation of marine traffic in an area from supervision. The macro situation of marine traffic describes the operation state of marine traffic system at macro level, reflecting the overall condition and evolution trend of the behavior characteristics and interaction of elements in marine traffic system.

In the actual navigation of ships, when there is no trend of convergence between ships, the distance between ships is still very short; however, when the complexity between ships greatly reduces, the complex relationship between two ships can be ignored. The convergence/divergence relationship between ships is judged by calculating the approaching rate when the complex network of marine traffic situation is constructed. It is to determine whether the two ships are connected or not. The approaching rate reflects the approaching effect of ships from time dimension. The degrees of convergence and divergence of ships are defined as the approaching rates between ships, which is expressed by the projection of the relative velocities of two ships at relative distances.

$$R_{ij}=\frac{\overrightarrow{D_{ij}}\bullet\overrightarrow{V_{ij}}}{\left\|\overrightarrow{D_{ij}}\right\|}=\left\|\overrightarrow{V_{ij}}\right\|\bullet\cos(\overrightarrow{D_{ij}},\overrightarrow{V_{ij}}) \quad （1）$$

where $\overrightarrow{D_{ij}},\overrightarrow{V_{ij}}$ are the relative distances and velocities of the two ships, respectively.

Figure 1 shows the construction process of complex marine traffic network. If the approaching rate is greater than or equal to 0, it means that the two ships are sailing in parallel or in a divergence trend. In such cases, the relationship between the two ships is not considered, and there is no connection between the two ships. If the approaching rate is less than 0, it indicates that there is a convergence trend between ships. At this time, the complexity between two ships is calculated as the weight of the edges in the complex network of marine traffic situation.

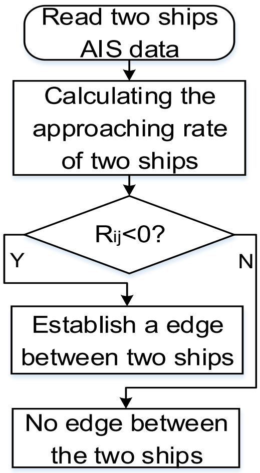

Read two ships AIS data
Calculating the approaching rate of two ships
Rij<0?
Y
N
Establish a edge between two ships
No edge between the two ships

Figure 1 Construction process of complex marine traffic network

2.3. Method of calculating edge weight

In our previous studies on the complexity of marine traffic flow, ships are usually regarded as the most basic traffic unit. Velocity, course and position of ships are important parameters affecting the marine traffic complexity, and the complexity between ships can be calculated by them. In the work, the method of calculating the complexity of marine traffic flow in our previous study was used to calculate the complexity between two ships[6], and the complexity is mapped to the weight of the edges in the complex network of marine traffic situation.

Figure 2 shows three marine traffic situations with different structures. The red node represents ship; edge the relationship between ships; w the edge weight. In Figure 2(a), three ships under the situation shows a disperse trend, so they do not influence each other. In Figure 2(b), three ships shows a converge trend, without influencing each other in different ways. At this time, the situation is more complex.

As shown in Figure 2(c), there are only two ships in a converge trend, to which controllers should pay more attention. The between-ship complexity in Figure 2(c) is relatively low. Although there is the same distance between ships and the same density in the three traffic situations in Figure 2, the complexity for controller and mariner differs. In the current study, the analysis of structure in marine traffic situation is highly deficient.

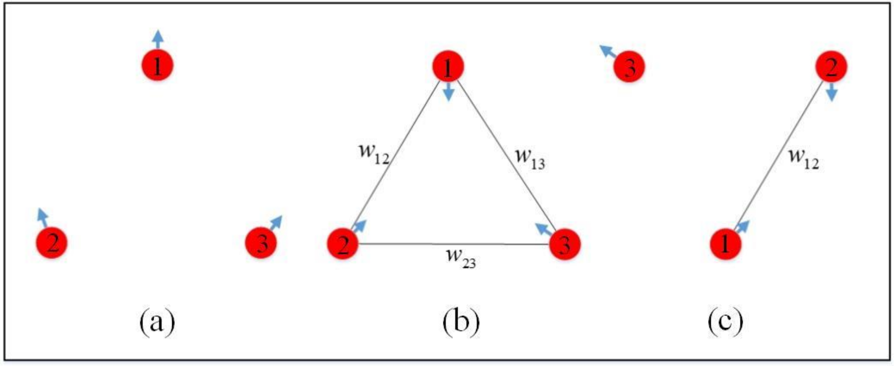


Figure 2 Different marine traffic situations from network structure

3. MARINE TRAFFIC SITUATION COMPLEX NETWORK TOPOLOGICAL CHARACTERISTICS

3.1. Research data

Taking the Zhenjiang Dagang section of the Yangtze River as the research object, the work selected the ship AIS data from 00:00-23:59 on 13 June, 2018 in this area to analyze the complex network of marine traffic situation. In order to reduce the computational complexity, the sampling interval was set to 10 minutes.

Firstly, we removed the data of ship length and width not conforming to the actual situation and ship position abnormality, speed abnormality, and course abnormality from experimental data set. Then the ship AIS data in research area was interpolated every second. Traffic separation scheme was adopted in Zhenjiang Dagang section of the Yangtze River. In order to analyze the differences of marine traffic situation between upstream and downstream traffic, the waters of Zhenjiang Dagagn were divided into the upstream and downstream areas.

$$A = \{Area1, Area2\}$$

Area 1 is the upstream area, and area 2 the downstream area (See Figure 3). Figure 4 shows the change of traffic volume in two areas.

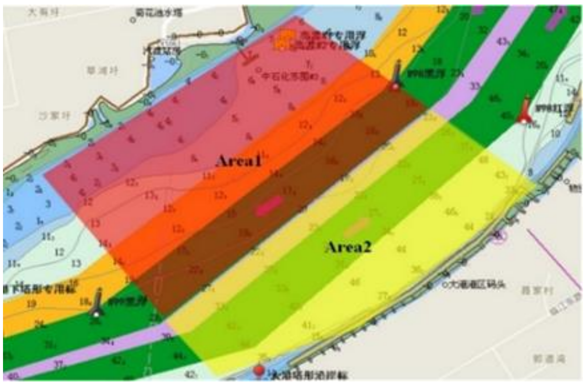


Figure 3 Upstream and downstream area in Zhenjiang Dagang waters

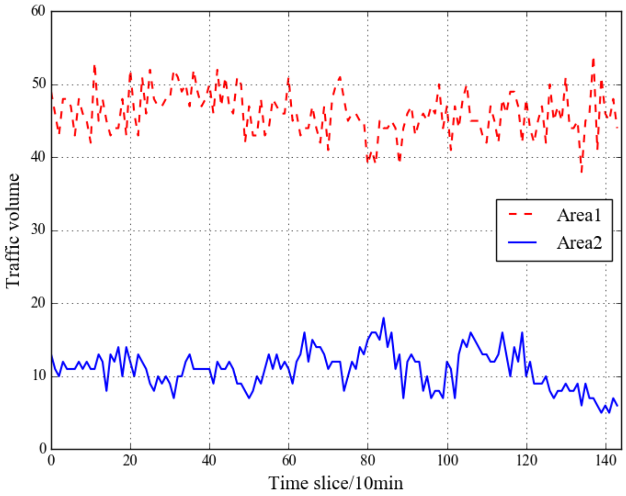


Figure 4 Traffic volume in areas 1 and 2

3.2. Complexity of marine traffic situation

The work was to propose an index to describe the complexity of marine traffic situation based on the topological characteristics of complex network. Firstly, two concepts are introduced—degree and vertex strength. Degree is a very important parameter in complex networks. N is defined as the number of nodes in a network, and the degree k_i of node n_i as the number of edges connected to node n_i. In the complex network of marine traffic situation, the degree of nodes represents the number of ships associated with a ship. The larger nodal degree means more ships having close relationship with the vessel, which causes the larger possibility of conflicts.

Figure 2(c) is a typical scenario, ships 2 and 3 are convergent, and ship 1 is on the contrary. Thus in Figure 2(c), $k_1 = 0$, and $k_2 = k_3 = 1$. In the complex network theory, the vertex strength is defined as the sum of edge weights associated with a node.

In most studies, the complexity of marine traffic situation is defined as the number of ship in unit area, and the larger number indicates the higher complexity. Although the method is simple and applicable, it cannot differentiate the marine traffic situation with the same number of ships.

The work used the vertex strength to evaluate the complexity of marine traffic situation in an area. The complexity was defined as the sum of vertex strength of all nodes in a marine traffic situation network, and reflected the average number of conflicts of nodes in the network. The concrete calculating methods are as follows.

$$S_i = \sum_{j=1}^{N} w_{ij} \tag{2}$$

$$U_i = \frac{S_i}{k_i} \tag{3}$$

$$M = \sum_{i=1}^{N} U_i \tag{4}$$

where S_i is the vertex strength of node n_i; w_{ij} the weight of the edge between nodes n_i and n_j; U_i the unit weight of node n_i; M the sum of all node unit weight of all nodes, reflecting the macro situation value in an area.

In the complexity of marine traffic situation in specified area was evaluated using the sum of unit weight of all nodes in a marine traffic situation network, reflecting the average number of conflicts of nodes in the network. In marine traffic, the larger number of ships indicates the smaller average distance between ship and the higher conflict probability. Figure 5 shows the curve of the situation complexity with time.

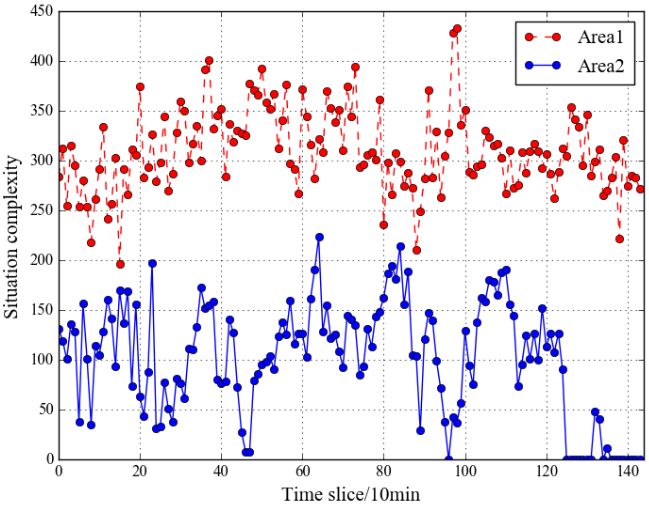


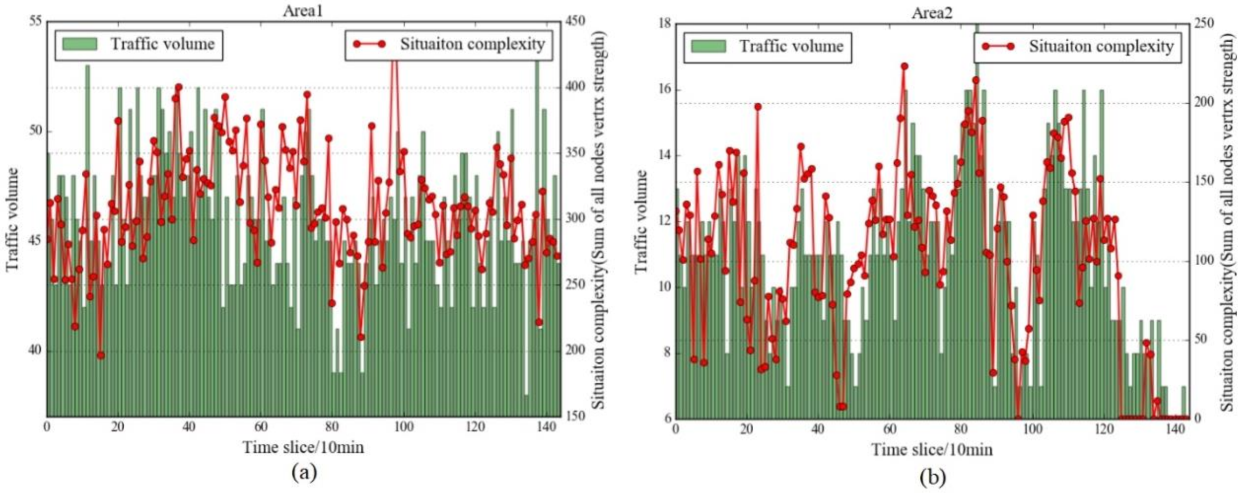
Figure 5 Change of the situation complexity in specified area.

In Figure 5, the situation complexity in area 1 is obviously more than that in area 2. In Figure 4, the traffic volume of area 1 is obviously more than that of area 2. There are similar change trend for the complexities of marine traffic situation in the two areas and the traffic volume, which can reflect the relationship between them. In general, the complexity of marine traffic situation increases with the number of ships, which proves the calculation model in the work.

However, there are differences between traffic situation complexity and traffic volume. In some cases, the influence of the number of ships on the complexity of marine traffic situation is different, and the situation complexity and traffic volume are not necessarily positively correlated. Figure 6 shows the relation between situation complexity and traffic volume.

In Figure 7, from the 9th to 12th time slice, four representative scenarios in area 1 are selected to validate the proposed model in the work. In four scenarios, the number of ship in area 1 at the 11th time slice is the largest, with the largest corresponding situation complexity. However, at the 10th time slice, the number of ship is the smallest, and the marine traffic situation complexity is not. Moreover, the number of ships at the 8th time slice is equal to that at the 12th, but the corresponding traffic situation is different. We analyze this phenomenon from the perspective of complex network.

Figure 8 shows the complex network of marine traffic situation corresponding to the four scenarios. In Figure 9, the number of complex network edges is not positively correlated with the number of ships, but positively correlated with the complexity of marine traffic situation. This indicates that the potential conflict of ships affects the marine traffic situation as well as the number of ships.



(a) (b)

Figure 6 Relation between situation complexity and traffic volume

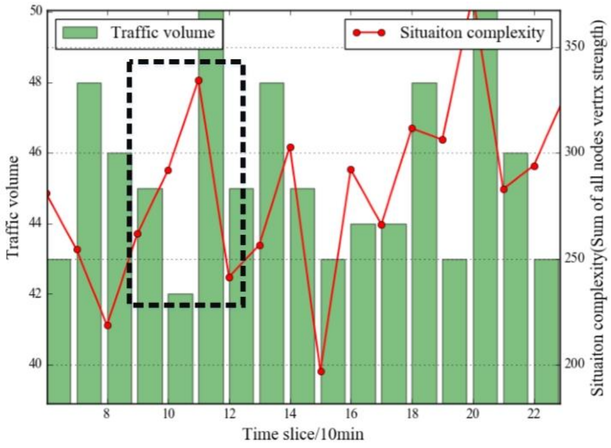
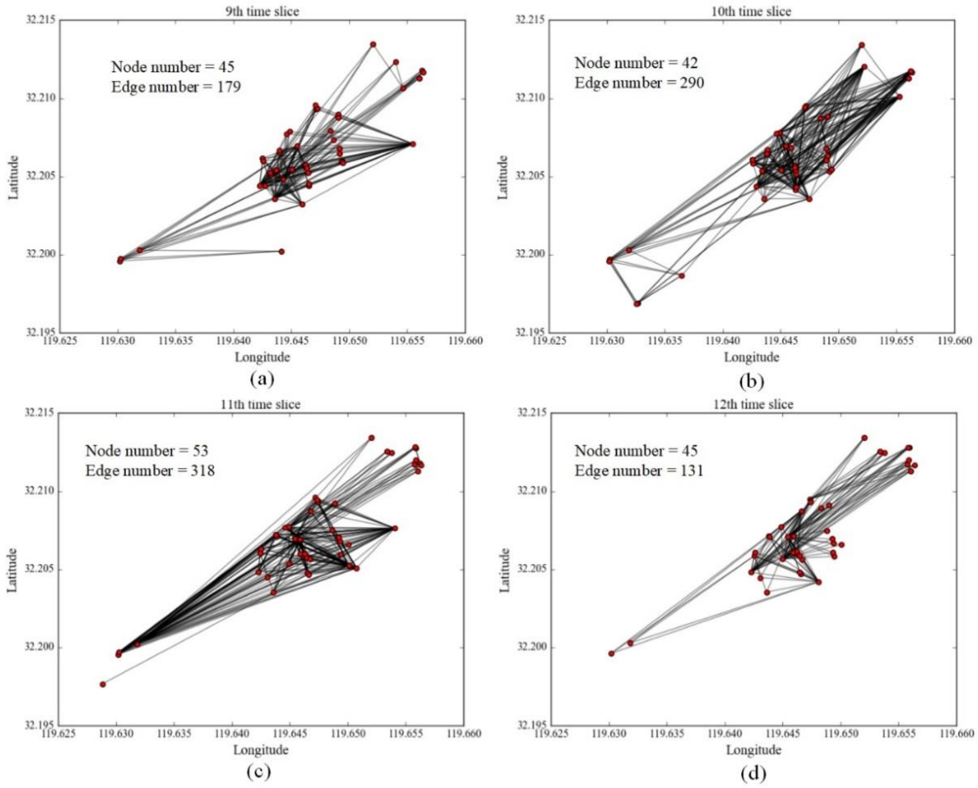


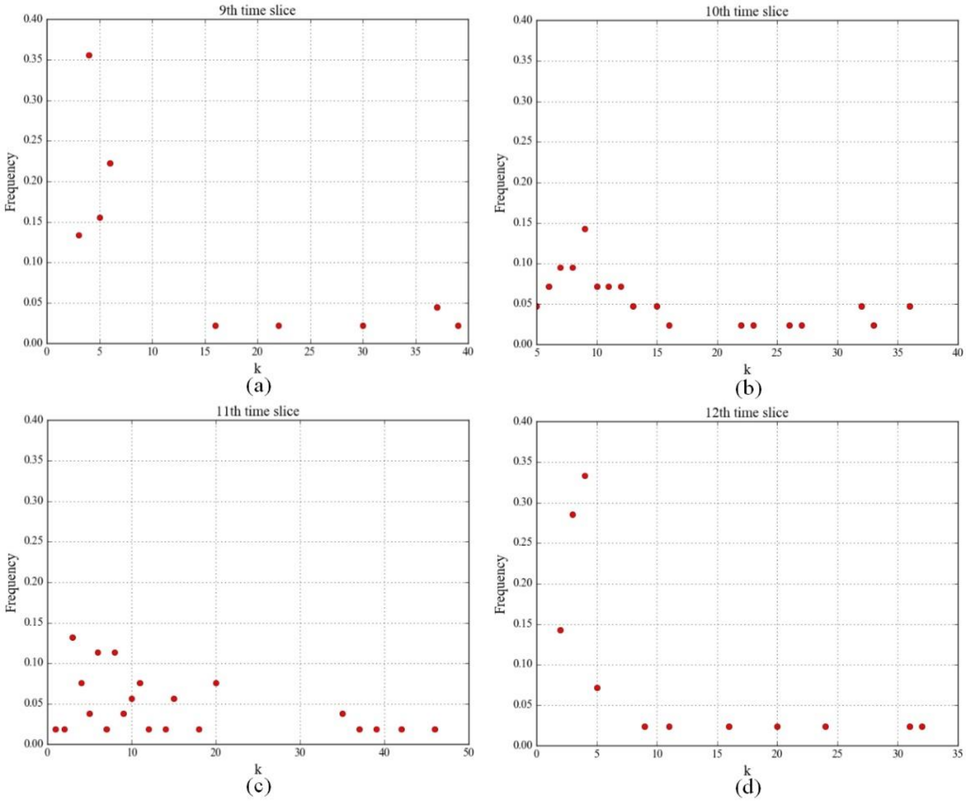
Figure 7 Actual scenarios selection in area1



(a) (b)

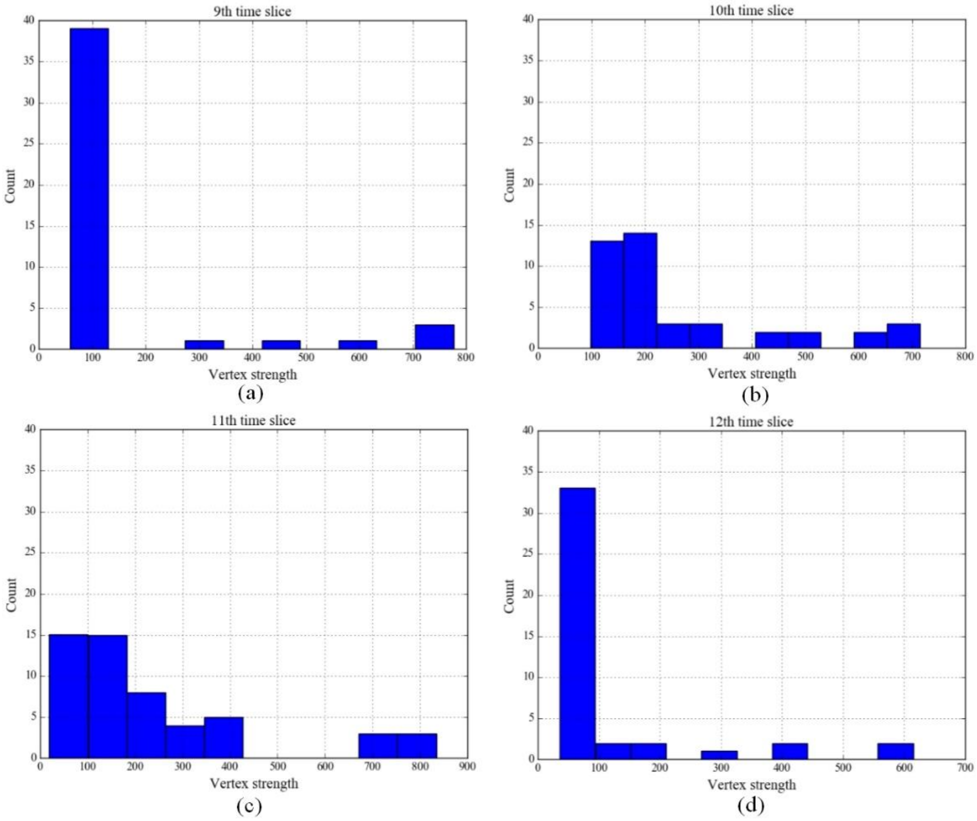
(c) (d)

Figure 8 Network structure in four actual scenarios in area1



(a) (b) (c) (d)

Figure 9 Degree distribution in four actual scenarios in area1



(a) (b) (c) (d)

Figure10 Vertex strength distribution in four actual scenarios in area 1

Figure 9 shows the degree distribution in four actual scenarios in area 1. When the situation complexity is low, the degree distribution of the network is in the power law distribution (See Figures 9(a) and (d)). This is because when the complexity of the area is low, most of the ships have fewer conflicts, and only a few ships are crucial. These crucial ships are the key nodes in the complex networks of marine traffic situation, while other ships are peripheral nodes.

When the situation complexity is high, the complex network of marine traffic situation has a uniform degree distribution (See Figures 9(b) and (c)). For the same reason, when the complexity of marine traffic situation increases, most ships are caught in a complex conflict situation, and the degree distribution of the complex network is more uniform. The distribution of vertex strength in Figure11 has the same trend with degree distribution.

Four scenarios were analyzed to verify the rationality of the proposed model. The results of the work are in line with the actual situation. It can be concluded that traffic volume only represents the number of ships in specific area, rather than the relationship between ships. The macro situation complexity model proposed in the work reflects the change of complexity of marine traffic situation sensitively, and provides the theoretical support for marine traffic management.

3.3. Homogeneity of marine traffic situation

In the marine traffic system, the high density of ships in an area leads to the high probability of conflicts. Some ships are in low traffic density area, and the probability of conflict is low in the short term. In other words, different ships have different impacts on the marine traffic system because of their different motion states and traffic densities. It conforms to the essence of the marine traffic system as a non-linear system. From the perspective of complex networks, this phenomenon reflects the importance of different nodes in complex networks. Therefore, the work put forward the homogeneity of marine traffic situation, and the degree of homogeneity affected by ships in the region. The higher homogeneity of the situation means more uniform the marine traffic situation is affected by ships. The structure entropy of complex network is introduced as a macro index to measure the homogeneity of marine traffic situation. It can accurately reflect the structural differences of marine traffic situation in different time.

For the power law degree distribution of marine traffic situation networks, there are core nodes with a large number of connections and most peripheral nodes with a small number of connections (See Figures 8(a) and (d)). The importance of each node is different, which reflects the complex network of marine traffic situation is a non-homogeneous. In the complex network of marine traffic situation, the importance of nodes is defined as

$$I_i = \frac{k_i}{\sum_{j=1}^{N} k_j} \qquad （6）$$

Without considering k=0, the network structure entropy is defined as

$$E = -\sum_{i=1}^{N} I_i \bullet \ln I_i \qquad （7）$$

Figure 11 shows the change of the network structure entropy in areas 1 and 2. With a larger number of ships, the influence degrees of ships on the overall marine traffic situation are homogeneous. The structure entropy in area 1 is higher than that in area 2. Figure 12 shows the relation between structure entropy and situation complexity, and network structure entropy and situation complexity have strong correlation. To explain this phenomenon, four scenarios in Section 3.2 are selected (See Figure 13).

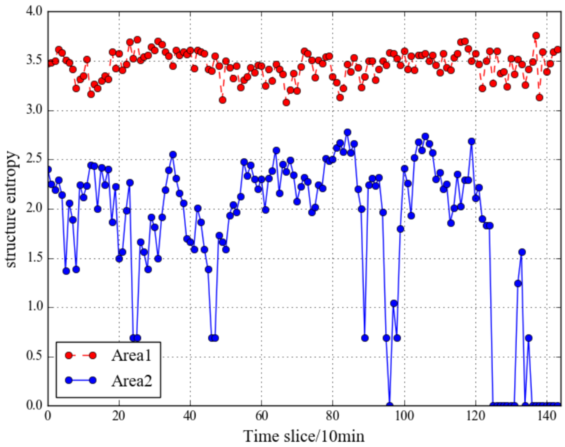
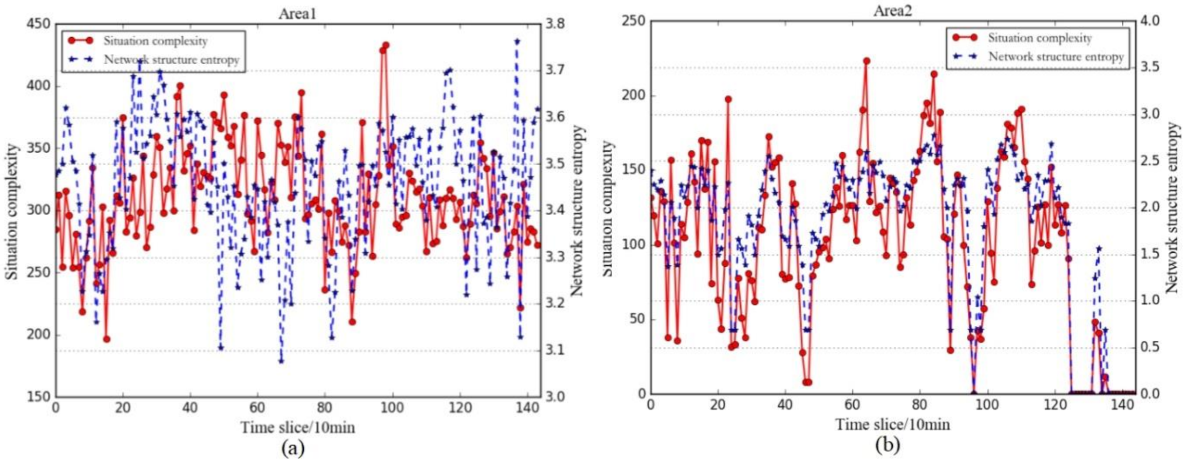


Figure 11 Change of the network entropy in areas 1 and 2



(a) (b)

Figure 12 Relation between structure entropy and situation complexity

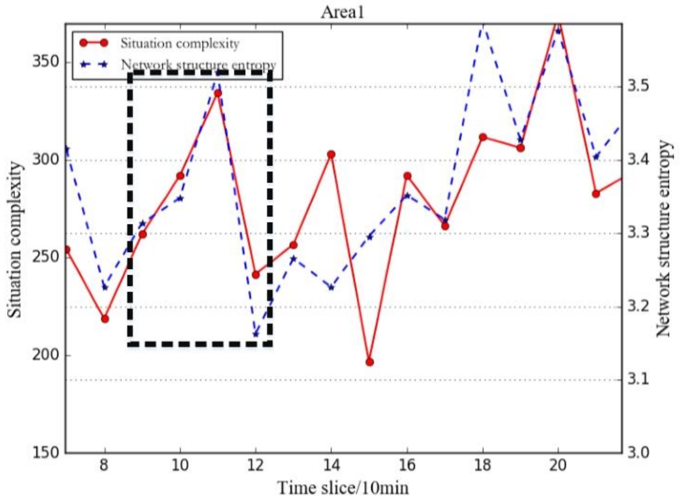
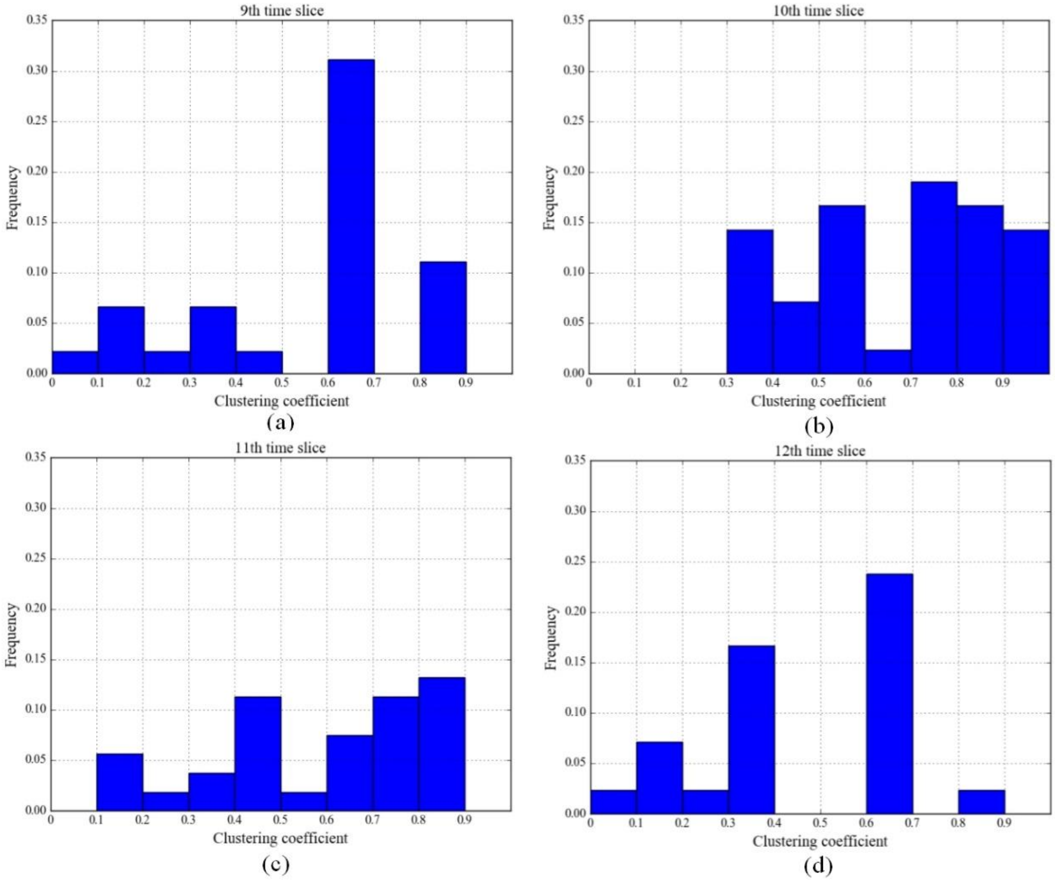


Figure 13 Actual scenarios selection in area 1

The complex networks at the 9th and 12th time slices in area 1 only have few core nodes, so the situation complexity (non-homogeneous) of area 1 is greatly affected by a few ships. This phenomenon can be explained by the clustering coefficient of the node in network. The definition of clustering coefficient is as follows.

$$C_i = \frac{E_i}{C_{k_i}^2} \tag{8}$$

where C_i is the clustering coefficient of node v_i; E_i the actual number of edges in v_i's k_i neighbour nodes; $C_{k_i}^2$ the number of all possible edges.



(a) (b) (c) (d)

Figure 14 Clustering coefficient distribution in four actual scenarios in area 1

The clustering coefficient distribution in Figures 14(a) and (d) is obviously uneven. The proportion of nodes is smaller with higher clustering coefficient. The clustering coefficient distribution in Figures 14(b) and (c) is uniform, and the proportion of high clustering coefficients is large. This proves that most nodes are in a highly connected situation, and the complexity of marine traffic situation is non-homogeneous. In practical application, the homogeneity of situation can sensitively reflect the differences of situation structure of marine traffic system. Managers can quickly form situational awareness according to different situations and take corresponding management countermeasures.

4. CONCLUSION

In the work, the complex network theory was introduced to map the complex relationship between multiple ships into the complex network, with the complex network of marine traffic situation established. The sum of unit weights of all nodes was used to indicate the complexity of marine traffic situation. The concept of homogeneity of marine traffic situation was put forward, with the homogeneity of marine traffic situation reflected by network structure entropy. The composition of marine traffic system and the evolution of system state were explained in principle. Finally, using ship AIS data in Zhenjiang Dagang waters, four examples were given to analyze the complex network model of marine traffic situation. The model of marine traffic situation based on complex network considered the non-linear influence among multiple vessels rather than simply superimposing the situation values of single ship. It highlights the structural characteristics of marine traffic system, providing a new idea for expressing the situation of marine traffic.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers and editors for their constructive comments, which is very helpful to improve the paper. This work was supported by the National Science Foundation of China (NSFC) (Grant No. 51679180), Double First-rate Project of WUT, the National Key R&D Program of China (Grant No. 2018YFC1407405), and the National Key R&D Program of China (Grant No. 2018YFC0213900).

REFERENCES

[1] Zissis, D., Xidias, E. K., & Lekkas, D. (2015). A cloud based architecture capable of perceiving and predicting multiple vessel behaviour. Applied Soft Computing, 35, 652-661.

[2] Zhang, S. K. , Shi, G. Y. , Liu, Z. J. , Zhao, Z. W., &Wu, Z. L. . (2018). Data-driven based automatic maritime routing from massive ais trajectories in the face of disparity. Ocean Engineering, 155, 240-250.

[3] Mou, J. M., Tak, C. V. D. , &Ligteringen, H. (2010). Study on collision avoidance in busy waterways by using ais data. Ocean Engineering, 37(5-6), 483-490.

[4] Zhang, W. , Goerlandt, F. , Kujala, P., & Wang, Y. . (2016). An advanced method for detecting possible near miss ship collisions from ais data. Ocean Engineering, 124, 141-156.

[5] Zhang, W. , Goerlandt, F. , Montewka, J. , & Kujala, P. . (2015). A method for detecting possible near miss ship collisions from ais data. Ocean Engineering, 107, 60-69.

[6] Wen, Y. , Huang, Y. , Zhou, C. , Yang, J. , Xiao, C. , & Wu, X. . (2015). Modelling of marine traffic flow complexity. Ocean Engineering, 104, 500-510.

[7] Kang, L., Meng, Q. , Liu, Q. . (2018). Fundamental diagram of ship traffic in the Singapore strait. Ocean Engineering, 147:340-354.

[8] Li, L., Lu, W., Niu, J., Liu, J., & Liu, D. (2018). AIS Data-based Decision Model for Navigation Risk in Sea Areas. Journal of Navigation, 71(3), 664-678.

[9] Hou, H. Q. , Li, Y. C. , He, W. , & Liu, X. L. . (2014). Vessel traffic flow distribution model of bridge area waterway in the middle stream of Yangtze River. Applied Mechanics and Materials, 551, 127-133.

[10] Zhang, L. , Meng, Q. , & Fwa, T. F. . (2017). Big ais data based spatial-temporal analyses of ship traffic in singapore port waters. Transportation Research Part E Logistics & Transportation Review.

[11] Altan, Y. C. , & Otay, E. N. . (2017). Maritime traffic analysis of the strait of istanbul based on ais data. Journal of Navigation, 1-16.

[12] Wu, X. , Mehta, A. L. , Zaloom, V. A. , & Craig, B. N. . (2016). Analysis of waterway transportation in southeast texas waterway based on ais data. Ocean Engineering, 121, 196-209.

[13] Balmat, J. F. , Frédéric Lafont, Maifret, R. , & Pessel, N. . (2009). Maritime risk assessment (marisa), a fuzzy approach to define an individual ship risk factor. Ocean Engineering, 36(15-16), 1278-1286.

[14] Weng, J. , Meng, Q. , & Qu, X. . (2012). Vessel collision frequency estimation in the singapore strait. Journal of Navigation, 65(02), 207-221.

[15] Goerlandt, F. , & Kujala, P. . (2011). Traffic simulation based ship collision probability modeling. Reliability Engineering & System Safety, 96(1), 91-107.

[16] Goerlandt, F. , St?Hlberg, K. , & Kujala, P. . (2012). Influence of impact scenario models on collision risk analysis. Ocean Engineering, 47(none), 74-87.

[17] Huang, S. Y. , Hsu, W. J. , Fang, H. , & Song, T. . (2013). A marine traffic simulation system for hub ports. Acm Sigsim Conference on Principles of Advanced Discrete Simulation. ACM.

[18] Bomberger, N. A. , Rhodes, B. J. , Seibert, M. , & Waxman, A. M. . (2007). Associative Learning of Vessel Motion Patterns for Maritime Situation Awareness. 2006 9th International Conference on Information Fusion. IEEE.

[19] Riveiro, M. , Falkman, G. , & Ziemke, T. . (2008). Improving maritime anomaly detection and situation awareness through interactive visualization. Information Fusion, 2008 11th International Conference on. IEEE.

[20] Rhodes, B. J. , Bomberger, N. A. , & Zandipour, M. . (2007). Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness. International Conference on Information Fusion. IEEE.

[21] Yigit C. , Emre N. . (2018). Spatial mapping of encounter probability in congested waterways using AIS. Ocean Engineering, 164:26.3-271.

[22] Simsir, U. , & Ertugrul, S. . (2009). Prediction of manually controlled vessels' position and course

navigating in narrow waterways using Artificial Neural Networks. Elsevier Science Publishers B. V.

[23] Xiao, Z. , Ponnambalam, L. , Fu, X. , & Zhang, W. . (2017). Maritime traffic probabilistic forecasting based on vessels\" waterway patterns and motion behaviors. IEEE Transactions on Intelligent Transportation Systems, 1-13.

[24] Yan, Y. , Zhang, S. , Tang, J. , & Wang, X. . (2017). Understanding characteristics in multivariate traffic flow time series from complex network structure. Physica A: Statistical Mechanics and its Applications, 477, 149-160.

[25] Zanin, & Massimiliano. (2014). Network analysis reveals patterns behind air safety events. Physica A: Statistical Mechanics and its Applications, 401, 201-206.

[26] Ducruet, César, & Zaidi, F. . (2012). Maritime constellations: a complex network approach to shipping and ports. Maritime Policy & Management, 39(2), 151-168.

International Seminar on Safety and Security of Autonomous Vessels
17 - 18 September 2019, Helsinki

Trustworthy versus Explainable AI in Autonomous Vessels

Jon Arne Glomsrud[1,*], André Ødegårdstuen[2],
Asun Lera St. Clair[3] and Øyvind Smogeli[4]
[1,2,3,4] Digital Assurance Program, Group Technology and Research, DNV GL, Norway

ABSTRACT

Trust, the firm belief in the reliability, truth, or ability of someone or something, underlies all social and economic relations and is central for the acceptance and adoption of autonomous vessels both by the maritime community and the general public. Trust requires explanations but is a much broader concept facilitating interaction among people and between people and technologies.
Autonomous vessels are facilitated by artificial intelligence (AI), automating tasks previously performed by people, meaning that roles, responsibility, authority and decision making are delegated to data and algorithms. The need for trust, however, remains unchanged, but people become dependent as well as changed by these technologies. Simultaneously, multiple layers of interaction between people and technologies will likely continue to exist.
People need valid explanations and causal reasoning for trusting critical, surprising or unexpected behaviour or decision making, also in the context of autonomous vessels, as incorrect behaviours or decisions can quickly translate into critical consequences. Such trust also depends on technical assurance processes where we emphasize that explanations can and need to play a role as valid evidence. We argue that the current methods for explaining AI are insufficient in providing trust in autonomous vessels and are too narrowly framed towards developers of AI. With the multiple points of interaction between people and autonomy, we argue that it is urgent to identify and mature explanation methods suitable for all types of interactions during development, assurance or operation. Explanations should be adapted to roles and responsibilities, and aspects such as context, cognitive skills, alertness, contextual knowledge, and time available to act by the user. We propose four types of explanations to suit the *developer*, *assurance*, *end-user*, and *external* explanation needs, which must be mapped out before the design such that trustworthy, interacting autonomous vessels can emerge.

Keywords: Explainable AI; Autonomy; Trust; Human Machine Interaction, Assurance.

1. INTRODUCTION

Trust is central for the acceptance and adoption of autonomous vessels into the maritime domain. Trust underlies all social and economic relations and is the firm belief in the reliability, truth, or ability of someone or something. Trust requires explanations, but is a much broader concept, especially because trust facilitates interactions among people, and between people and technologies, an ability that makes trust such an asset. Trust in this context is not only the assurance of the technical safety or suitability of an autonomous vessel, but also the wider public trust and acceptance of such vessels.

AI plays a central role in autonomous vehicles; one can even consider AI and autonomy as synonymous given the deployment of AI in any transport system entails the transfer of decision making from humans to algorithms. This challenges the established technical trust mechanisms, especially the assurance mechanisms to which all vehicles are subjected to demonstrate their design and operations are fit for purpose. We define *assurance* as a structured collection of arguments supported by suitable *evidence* demonstrating that a system is fit for purpose. This

---

[*] Corresponding author: phone: +47 92403158, email: jon.arne.glomsrud@dnvgl.com

means that **from an assurance perspective, explanations need to become evidence**, and as such subject to stringent **standards of suitability**.

With the increasing interest in and use of AI, the need for explaining the decision-making performed by algorithms emerges as a key research field and is even implicitly and explicitly demanded in some regulations, e.g. the EU GDPR regulations makes references to "*the right ... to obtain an explanation of the decision reached*" (EU GDPR.ORG, 2018). Another key topic in the AI debate is the Explainable AI program (XAI) (DARPA, 2017). This program states that along with the success in Machine Learning, there is need for machines to explain their decisions and actions to human users.

So far, explainability of AI has primarily been focused on the researcher and developer of AI applications. There is little work asking what types of explainability will be required in the context of providing assurance. Also, there has been little attention to the role explainability can have in relation to a human user or to any other human agent affected by the deployment of AI in society, such as for example autonomous vehicles.

To enable trustworthy autonomous vehicles, it is therefore important to develop suitable explainability methods for these extended needs. In this paper we provide insights and examples to establish an extended notion of explainability within the context of the assurance, operation and interaction with AI enabled autonomous vessels, exemplified in the form of an autonomous small ferry.

The Paper is organized as follows: Section 2 establishes the background information around autonomous vehicles, AI and explainable AI, the role of trust, and the use-case of an autonomous ferry. Section 3 explores and discusses the role and need of explanations both in the context of the operation of an autonomous ferry (3.1), and in the context of assurance (3.2). Section 4 presents the main conclusions and discusses some lines of potential future research needs.

2. BACKGROUND

2.1. Autonomy and human interaction

We claim that autonomous vehicles too often are envisaged as disconnected from human interaction. However, it is in fact not possible to conceive any technology without human beings as creators or as users as well as impacted by them given autonomous vehicles increasingly are operating in public spaces. Autonomous vehicles always need to be designed to a certain operational design domain (ODD) (SAE International, 2018a), implicitly meaning there are certain built-in limitations to its abilities. If the operational situation exceeds the ODD or the ability is not sufficient, the vehicle must perform certain safety actions or handover control to human beings. Thus, multiple layers of interaction between human beings and machines will continue to exist.

Fully autonomous vehicles will most likely take some time and semi-autonomous concepts are at the present stage most relevant to consider. Several car manufacturers seem to experience concern and challenges related to Society of Automobile Engineers (SAE) level 3 of driving (SAE International, 2018b), where the car normally drives automatically, but the driver is expected to take over when requested by the automation. This handover is seen as hard to solve in a safe manner and puts high demands on the situational information presented to the driver, as well as the alertness and ability of the driver to immediately control the situation. It is not possible to envision the safety of a semi-autonomous system without looking into the requirements and conditions of the controller or co-agent of that autonomous system, in this case the driver. In fact, what autonomy is currently leading towards is distributed agency between humans and machines (Rammert, 2012). This case pinpoints the paradox that with increased automation follows the need for increased focus on human-machine interaction and co-behaviour. (Rahwan, et al., 2019) and (Parasuraman, Sheridan, & Wickens, 2000). (Parasuraman, Sheridan, & Wickens, 2000) have developed models for types and levels of human interaction with automation which in our view is suitable for a more objective analysis of human-machine interactions in the context of assurance. This model is also discussed and used further in a recently published guideline for the assurance of remote-controlled and autonomous ships (DNV GL, 2018).

A key issue with autonomy is that autonomous vehicles need to hand over control to humans at some point. At the instant an autonomous vehicle is not capable of handling an operative situation, there is need for handover to human control, regardless of whether this has been considered and designed. Similarly, for remotely controlled vehicles, in case of substantial loss of communication or when the remote-control capability is insufficient, an autonomous function will need to step in and control the situation (again, regardless of whether this has been considered and designed). Both these points can be summarized in a short conclusion; human intervention will always need to be a fallback for autonomous control, autonomous control will always need to be a fallback for human (remote) control, and this distributed agency or dual control requires trust between humans and machines.

2.2. AI and explainability in the context of autonomous vehicles

Presently, the advances in self-driving cars would not have been made possible especially without the breakthroughs in AI in the field of machine learning (ML), especially artificial neural networks (ANN) and computer vision technologies for sensing and analysing traffic situations. Such ML is trained on known and selected data-sets, called *data-driven* and *supervised* ML before being deployed into operation. AI has become a major enabler and a critical part of self-driving capabilities. The advances in AI based planning functions, e.g. in the field of reinforcement learning, make it possible that AI in the near future will become an even more central part of autonomous systems. AI in our context is mostly related to ML and applied to automate human tasks in operating a vehicle. We argue that it is important for the emerging field of explainable AI to explore explainability in the context of autonomous vehicles. For people interacting with autonomous vehicles, it does not matter what technology is used to achieve autonomy; the same human needs for explanation in the interactions are needed.

2.3. Explainable AI

The motivation for explainable AI (XAI) is multi-fold; detecting bias and spurious correlations and ensuring fairness and safety are some of the most frequently mentioned. In this paper, we assume that an explanation can be evaluated according to its *interpretability* and its *completeness* (Gilpin, et al., 2018).

There are in general two approaches in making models explainable: designing models to be explainable by nature or applying techniques for interpretation after the output (post-hoc). The explanations can be divided into two different types (Gilpin, et al., 2018). The first type is *explanations of processing* where one tracks inputs to outputs, e.g., by answering the question: "*Why does this particular input lead to that particular output?*". This can be considered more of a black box method that does not need access to the internals of the AI. The second type is *explanations of representation*, e.g., answering the question: "*What information does the network contain?*" (Gilpin, et al., 2018). This latter method needs some access to the internals of the AI and can be considered more of a white or grey box method. These different approaches and methods mean that the need for explainability should be mapped out before a system is designed, but this also means that the explanation is different for someone who is an end-user, a developer, or an *external* affected or forced to interact with an autonomous system (e.g. people external to the ferry, its passengers, operators or end-users). We assume that any explanation is directed towards a human user, regardless of what type.

With these assumptions in mind, we argue that the following should be taken into consideration during the mapping of explanations:

- *User needs*. To what end does the explanation serve? Lipton (Lipton, 2017) argues that AI professionals should be better at determining what the stakeholders want regarding interpretability, when for example, making appealing visualizations or claiming interpretability. This means one needs to understand who the users are as well as their needs.
- *Explanation strategies, cognitive bias and ability*. To meet the user needs, how interpretable and how complete does the explanation need to be? One should pay attention to making explanations user friendly, while controlling the risk of making them too persuasive (Herman, 2017) or entering conflict with our original goal of achieving understanding and trust. This balance will depend on the user's ability to understand the

model from which the explanation is generated. Thus, understanding user needs remains a fundamental first step.

- *Real-time vs. post-process*. Should the explanation be given in real time or generated retrospectively. Computational costs can limit the possible explanations in real-time.
- *Interpretability vs performance*. There might be trade-offs between interpretability and performance, e.g. a more complex model could perform better than a more interpretable model.

Considering that computer vision is an important part of autonomous vehicles and that modern computer vision models are convolutional neural networks, which are not explainable by design, we will assume that explanations for autonomous vessels to a large degree will consist of interpreting the decisions made by or based on ANNs. Various techniques for interpretations have been developed, such as Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju, et al., 2017), LIME (Ribeiro, Singh, & Guestrin, 2016) and SHAP (Lundberg & Lee, 2017). Such techniques can help us getting insight into how the ANN works and detect spurious correlations, which can be useful for a developer, but perhaps not so much for an end-user.

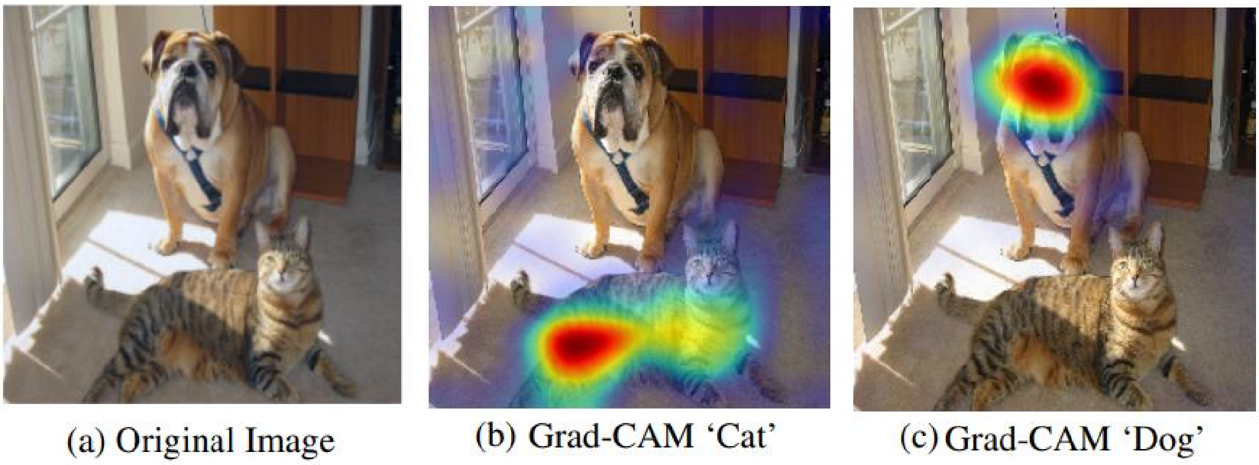

(a) Original Image (b) Grad-CAM 'Cat' (c) Grad-CAM 'Dog'

Figure 1: Results from applying Grad-CAM (Selvaraju, et al., 2017) to an ANN classifier for cats and dogs. This technique uses information about the model weights to illustrate which regions of the image that contributed to the decision.

It can be worth considering techniques that do not attempt to interpret the models, but rather act as an interface between the model and the end-user in real applications. In the use case described in 2.5 and discussed in 3.1, the user responsible for the remote monitoring will benefit from receiving explanations of predictions of the operative environment and the planned behaviour made by the autonomous ferry. This can allow the operator to correct the ferry in case of wrong predictions or decisions. Such methods may be outside the scope of the original XAI program (DARPA, 2017), but can give valuable explanations in the context of autonomous vehicles.

2.4. Trust

Trust is central for the acceptance and adoption of any technology, and it has both an intrinsic and an instrumental value. Trust is the firm belief in the reliability, truth, or ability of someone or something. As such, it has intrinsic value in any context as it underlies all social and economic relations. Clearly, for autonomous vessels to be deployed, a wide societal trust in those vessels is needed, otherwise their deployment will be compromised. But trust also has a critical instrumental value, as it acts as a facilitator of interactions among people. Consider for example the role trust has in contractual obligations: if all parties trust that the other parties will meet their duties, this prevents unnecessary controls and overhead in all steps of the contractual process. It is this second type of value, the instrumental value, that is of critical importance in the context of autonomous vessels, which are complex cyber-physical systems formed by interactions between people and technology.

The facilitation role of trust must go beyond trust among people, to include trust between people and technologies, facilitating relations among the members of a system. In the case of autonomous vessels, the system is formed both by people, technologies, and by artificial agents, algorithms with the capability to make decisions. Taddeo names these complex, intelligent systems as a hybrid system (Taddeo, 2017). When we delegate to digital technologies cognitive tasks that were earlier performed by humans, trusting them may be seen as a question of three dimensions that emerge in the interface between intelligent technologies and trust (Taddeo, 2017):

- General trust in the nature of technology,
- Trust in digital environments,
- The relation between trust, technology and design.

Taddeo's categorization fits well with autonomous vessels which are formed by different kinds of technologies, including intelligent ones. Explainable AI will be one important element in this system but will cover only the third sense of trust, in relation to technological design.

As outlined in 2.3, we evaluate explainability according to its interpretability and its completeness. Bias, transparency, robustness, reliability, lineage, trust in data and trust in models are at the centre of attention within explainable AI. Most of these issues are questions related to the trust and facilitation role of trust in the processes of technological design. The expected end result is an AI application that is transparent, able to be understood by humans, and able to explain how it has made a decision or prediction. However, *trusting* an autonomous system is something different, with more layers of complexity and of a broader nature. Yes, we will need to have trust in the data, trust in the models that generate predictions, but one will need to trust not only the algorithms but also the contexts in which this algorithm operates (both technical and non-technical). This relates to the importance of user needs, and knowledge of how much interpretability is sufficient in specific contexts. In addition, we need a clear understanding of the impact the AI in question may have on the autonomy of human users. As we have indicated in 2.1, increased automation leads paradoxically to an increased attention to human-machine interactions and co-behaviour, giving changes in the behaviour of the system affecting humans, and changes in human behaviour affecting machines (Rahwan, et al., 2019).

In short, to build trustworthy and explainable autonomous vessels, a perspective that looks only at the technologies themselves, no matter how explainable those technologies are, is insufficient. We need to understand the context or wider system in which autonomous vessels are deployed. This wider system may include users' perceptions and expectations, other agents, actors, structures, and relevant rules and regulations. Making the AI deployed in autonomous vessels explainable is important, but not enough to make it trustworthy.

Trustworthiness is also at the core of assurance. We often refer to third-party verification and certification as 'assurance'. Assurance refers to the structured collection of arguments supported by validation of suitable evidence which provides the confidence that a product or process is fit for purpose, and that it complies with safety, environmental, or other technical requirements. The provision of assurance is always based on credible technical information or knowledge, often validated by independent actors, to comply with existing regulations.

2.5. The Case of an autonomous ferry

This paper makes use of a conceptual autonomous short distance passenger ferry case for further concept developments and discussions. Such small ferry concepts are low-cost alternatives for bridges, tunnels and manned ferries. Several different concepts are currently being developed. For example, cross disciplinary research has been established in Norway (NTNU, 2019). The concept in this paper is a small unmanned ferry capable of autonomously onboarding passengers, undocking, manoeuvring and navigating to another quay, docking and finally offboarding the passengers. Surrounding traffic is avoided by inclusion of a collision avoidance system. The ferry must be remotely monitored for halt of operation, maintenance purposes, etc. The concept is interesting, having many relevant end-users and different human interaction points, e.g. development, operation, security and maintenance, maritime traffic in the vicinity of the ferry, and not the least, the passengers in normal or even critical situations.

Figure 2: One vision of what the next generation of driverless ferries may look like. Credit: Illustration: Reaktor (Finland - www.reaktor.com)

3. ROLE AND NEED OF EXPLANATIONS

3.1. Situational explanation needs

Explanations are a definite human need, but we might envision that also machines in the future can benefit from explanations, especially when explaining the future behaviour of interacting agents. When something in a situation or a context is important, critical, surprising, unexpected, or interesting, humans want explanations to understand, learn, and accept. Explanations are a guarantee for trust. Explanation needs often depend on the role, responsibility or consequences of something on people in a particular situation. The human ability to understand an explanation, i.e. to which degree an explanation is interpretable, is affected by the cognitive ability, alertness, contextual and tacit knowledge, the time available for interpreting the explanation, and of course the interpretability of the explanation itself.

Returning to our use case of an autonomous ferry, we can find different points of human interaction and need for explanations within the development and operation of the systems:

For the *developer* to understand or learn, or verify, improve or make the AI or autonomous system comply to requirements. This will typically require explanations that are relatively complete but harder to interpret. A truly complete understanding of the AI models can be out of reach, but by using best practices a developer should get a sufficient intuition on how the models work and what their weaknesses are. A full discussion on explanations in an assurance context is found in 3.2.

For *externals* like swimmers, kayakers, boats and vessels that are close to and interacting with the ferry to understand its intentions as early as possible. With autonomous and unmanned vehicles, the human to human communication that can inform if the vehicle has seen the swimmer or other vehicle and intends to act safely is lost and needs to be replaced with something new. An interesting example is the smiling car concept (Semcon, 2019), that tries to explain the intentions (stopping) of the car to detected pedestrians by showing a smile in the car front display to reassure that the pedestrians have been detected and the car intends to stop.

For *remote operators* monitoring the operation to obtain sufficient situational awareness and ability to predict the vessel's behaviour in time to intervene if needed. This could include communicating to the operator what situational elements matters for the situational understanding, the current chosen planned path, the estimated and predicted path of other vessels, and more conventional information like the health status of critical system components. This is comparable to the driver handover situation for self-driving cars at SAE Level 3. The ferry concept developers need a well-considered safety philosophy and careful consideration of the level of remote control or intervention needed both in normal and abnormal situations (DNV GL, 2018). The human intervention challenge is far from new and has generally followed the increased use of automation for decades (Parasuraman, Sheridan, & Wickens, 2000). We emphasize the need to carefully consider what information that explains the current decisions and status of the autonomous ferry effectively, such that the human remote operator can trust its ability to bring passengers safely across the water.

For *passengers* in a variety of journey phases or possible critical situations: (i) After *boarding* when the passengers have boarded and are waiting for undocking and start of crossing, the passengers would feel reassured to receive a signal or confirmation that the onboarding is safely finished, and the undocking/crossing can start. (ii) During *crossing* when the ferry navigates close to swimmers, kayakers, boats or vessels, or possibly large wildlife or other objects, passengers would want to know the intentions of the ferry. Is the ferry heading forward or planning to yield and let the traffic pass? Like the discussion for externals and remote operators above, a simple message could explain if the object is detected and if the ferry intends to yield or continue. (iii) During *approach* when the ferry approaches the destination quay and starts the docking phase. Like the situation with navigating close to objects or vessels, passengers would be reassured to receive a simple message that the quay is found, positioned and the ferry starts its docking procedure to safely dock. (iv) And finally in *abnormal situations* when the ferry experiences an abnormal situation, e.g. a critical system failure, unusual or unexpected environmental conditions or non-conforming nearby vessels, passengers will want to be informed even if this is not an obvious emergency. Normally, the safest place for passengers is onboard, unless the ferry is in danger of colliding, on fire, sinking, etc. Yet passengers will always want to know the exact situation. Explanations with the intention of informing what is going on should prevent panic and rather convey security and reassurance. Of course, this requires that the ferry or the remote operator detects the non-normal condition, and if this is not the case, it may be necessary to have some way for passengers to intervene and take control of the ferry, like an emergency stop button in an elevator. Such a situation is nevertheless not part of the field of explainable autonomy.

One can argue that using such a ferry is nothing more than an advanced elevator, but we are so familiar with elevators that we know e.g. that when the doors close, the elevator will soon move, that the elevator will not move until the doors close, that the doors will not open while moving between floors, and that the elevator will not fall down. The elevator industry is mature and have trustworthy systems, arrangements, regulations and organizations, but this maturity has developed over time. In addition, an elevator has an alarm button, presumably enabling those being transported to reach a human operator in case of an emergency. Time always leads to risk mitigation and risk acceptance, from which trust emerges. A new concept like autonomous ferries will not instantly earn the same level of trust as elevators, thus richer sets of explanations to assure the passengers that the ferry is controlling the steps and phases of the full journey correctly and safely are needed.

The above discussion of explanation needs gives us four different types of human interaction and associated explanations:

Developer explanations are for the researcher or developer on how the AI based systems work to understand and learn, or verify, improve or make the system comply to requirements. We claim that this has been the primary use of explainability so far. It is of course critical that the developer can trust that the systems work as intended and we support the developments in XAI towards this goal.

Assurance explanations are used in an assurance context, which will be discussed in 3.2 and which should be an important gap to fill for explainability of AI to become closer to the notion of Trustworthy AI. Assurance explanations need to be suitable as evidence in an external and independent assessment.

End-user explanations for end-users in an operative situation. This is not really a new problem, but rather one that generally has been around for decades. The new challenge is that the emerging AI technology is increasingly less interpretable or explainable than conventional software and used in increasingly more autonomous operative settings. We are unfamiliar both with the technology and the contexts of the human-machine interaction. This is where the differences in user training and capabilities do matter; The end-users can be divided into different categories, some familiar and trained to interact or operate the system, like the remote operator in the use case above, but some not trained or familiar with the system at all, like the ferry passengers. These differences in the knowledge or cognitive skills pose a challenge when designing the systems to interact safely and securely with the different categories of end-users.

External explanations for externals to the ferry, its passengers, operators or end-users. One can argue these share traits with e.g. the ferry passengers, but we define them as a separate type. It is important to realize that one cannot expect externals to know or understand that the ferry is autonomous, compared to ferry passengers who will realize this either as soon as they enter the ferry or during the journey. We believe this is an important aspect differentiating explanations for end-users and externals.

Common for *developer* and *assurance* explanations is that the explanatory situations are not in real operating time, rather the explanations can be produced without tight time constraints. As discussed in 2.3, the explanation does not need to be produced in real-time, and post-processing is enough. Nevertheless, the computational cost (time) to produce the explanations is still a limiting factor.

Common for *end-user* and *external* explanations is that it is crucial to analyse the real-time interaction situations thoroughly and evaluate them based on aspects mentioned earlier, such as the end-user cognitive ability, the alertness and the time available to understand and act. One can base these analyses on work like (Parasuraman, Sheridan, & Wickens, 2000) and (DNV GL, 2018), but with the increasing use of autonomous systems in hybrid human-machine interaction contexts, we may realize this is a new and unexplored field (Rahwan, et al., 2019).

3.2. Explanation needs in an assurance context

In 3.1, *assurance explanations* were intended as evidence in assurance and will be discussed below. *Assurance* is a structured collection of arguments supported by suitable *evidence* demonstrating that a system is fit for purpose. In practice, assurance is often the systematic collection of evidence from two aspects of the system and its development, that is firstly evidence that the requirements for the development of a system are complete and relevant, both requirements for the development process as well as the system itself, and secondly evidence that the development process and the developed and operated system is according to these requirements. The evidence is collected and documented from various activities, using suitable tools and methods and with needed participants and roles with the proper independence from the development. Evidence is specific to these activities, tools and methods and can in general be quite diverse. When the collected evidence is considered valid and complete, the assured system is also considered trustworthy, i.e. creating the necessary trust. In the context of explainability, it is important to note that not all explanations have the properties to be considered valid evidence. We can therefore envisage that explanation of AI or autonomous systems can support the assurance process, only if explanations are sufficiently valid and suitable to become evidence. There is a need for further research beyond this paper on methods of explainability and their individual suitability as evidence in an assurance process.

With data-driven AI, like supervised ML, a data set is split and used to train and test a model respectively. The data and the testing is therefore at the core of the AI development. As we discussed in 2.3, explanation types were in general explanations of the *processing* or of the *representation* in the AI model (Gilpin, et al., 2018). For software in general, two types of testing are commonly used, being *white box* and *black box* testing. White box testing refers to software code review and analysis, requiring access to the source code, whereas black box testing uses the executable software code, requiring only access to its inputs and outputs. In the latter the software source code can be kept confidential. AI is software even though not as readable as conventional source code, and the same concepts of testing could be applied. Explanations of processing may be more relevantly used as black box testing methods since it does not need access to the internals of the AI. However, explanations of representation may be viewed more as a white box testing method where inner workings of the AI are explained. Any explanation method aims to explain the AI models, the developed *product* and not the applied development *process* and we therefore argue that such explanations are mostly suitable as evidence related to the testing of AI models (both black and white box testing) and less relevant as evidence for quality or suitability of the development process.

The current methods for generating explanations of processing for ANNs, as mentioned in 2.4, are often limited to analysing how different parts of the inputs affect the activation of the individual layers in the ANN model, or more often the output of it. While this can give us some intuition on how the model behaves, it gives limited insights to the inner workings of it. Common types of ANNs are those who categorize (or *classify*) images based on their contents. Generating a map of areas in the image important for the model classification is a common type of explanation generated for this kind of network, and an explanation typical for explanations of *processing*. Generating a map that a user agrees with does not tell why the model identifies specific areas as important or if the user is likely to agree for other images. This unknown degree of generalization is a well-known, general problem for ANNs that extends to the explanations that can be generated from them. In an assurance context this restricts the use of explanations. We argue that because of the potential lack of generalization, explanations of processing must be used with a statistical approach to build confidence in the explanations.

The receiver of an explanation (the *explainee*) in an assurance context might have less knowledge and understanding of AI than a researcher or developer of AI, and his or her cognitive ability is thus lower. Therefore, assurance explanations need to be more interpretable than developer explanations. Simultaneously, the explanation needs to be highly correct such that it becomes valid assurance evidence. AI typically solves problems that are too complex to be solved with conventional software, so we assume that the AI solutions become complex. In order to be interpretable for humans, explanations need to reduce this complexity, but in doing so they risk removing important aspects of why the AI work and are hence less valid. Therefore, interpretability and validity (completeness) of explanations can be contradicting, and a potential challenge for assurance explanations. When considering if the explanation targets explanation of processing (black box testing) or explanation of representation (white box testing), we claim that explanations of processing are more interpretable by nature than explanations of representation, given that in general with black box methods, an understanding or knowledge of the inner workings of the black box is not needed. Still, explanation of processing often produces a separate simplified model to explain the black box and this model must of course be interpretable for the assurance explainee. In conclusion, as a general observation for explainability evidence for AI, the cognitive ability of the explainee is important to consider in order to ensure the minimum interpretability and efficiency of evidence. As a closing comment, evidence in an assurance context need to be sufficient, complete (e.g. statistically sound), valid and convincing. Even if explanations are interpretable and correct, they are not necessarily sufficient and complete in an assurance context.

Explainable AI is often discussed in relation to trust in AI, but as we have argued earlier, explanations are only a subset of trust. We believe it is more important to generate trust in the wider context of the operations of the autonomous vessels and not only in specific technical systems or solutions. Assurance of AI-based systems must instead of explanations be based primarily on testing. While conventional testing of systems aims to test over the range of possible scenarios or near the boundary conditions, AI-based systems are often implemented in such a way that the possible number of scenarios is near infinite, and with boundary conditions that are difficult to define in a high-dimensional space. This requires a different approach to testing. Intel, in collaboration with 10 industry leaders in automotive and autonomous driving technology, have published a framework (Wood, et al., 2019) for the design, development, verification and validation of safe automated passenger vehicles. This framework suggests a broader testing regime, including a statistical approach in a real-world setting, scenario-based testing, lifetime field monitoring and simulation. While this approach is time consuming and most likely expensive, we believe it is applicable and required for autonomous vessels.

4. CONCLUSIONS AND FUTURE WORK

In this paper we have argued that the topic of explainable AI needs to be unpacked in relation to both the users for whom the explanations are needed, and the different types of explanations required. We argue that explanation of processing and explanation of representation are of different natures and play different roles. In addition, we argue that trustworthy AI in the context of autonomy is a broader concept than explainable AI. Trustworthiness depends on that the different users and the different types of explanations have been successfully matched. That is, it is not an a priori but rather a posteriori issue.

Human-machine interaction is a key feature of autonomous systems and a key topic to be addressed when exploring explainability in the context of autonomous vehicles. Interpretability and explainability methods are a part of this interaction but have so far focused on the research and development of AI. We argue there is a need for developing suitable methods for the assurance process, the end-users, and externals interacting or being affected by the AI. In this paper we have three specific users in mind: the users of the ferry, externals affected by the autonomous ferry, and assurance participants.

We have proposed different types of explanations and briefly discussed the needs for each type. Generally, end-users need real-time explanations that are interpretable and adapted to their cognitive abilities, alertness and available time to understand and act. In both an assurance context and during development, explanations should be as complete and valid as possible, and simultaneously they must be interpretable to the explainee. We argue that an understanding of the required level of interpretability and completeness of evidence is needed prior to the actual development of an autonomous system.

Explanation of processing, which maps inputs to outputs and treats the AI as a black box model, may be considered more relevant for assurance and more explainable by nature than explanation of representation, which treats the AI as a grey or white box model. However, the challenge in assurance is not the interpretability or explainability itself, but rather if the set of explanations combined can suffice as valid, complete and convincing assurance evidence. We have stated that this could be possible but will require dedicated methods and practices to be developed.

In order to be interpretable, explanations are often simplified compared to what they try to explain. This means they can be less correct or complete than the actual system and should be used with precaution in any situation requiring trust in the AI, either towards an end-user or in an assurance context.

ACKNOWLEDGEMENTS

We acknowledge that our views are the result of ongoing cooperation with NTNU on research in the fields of autonomous vessels and AI and the excellent autonomous ferry concept use-case (NTNU, 2019).

REFERENCES

DARPA. (2017, May 01). *Explainable Artificial Intelligence (XAI)*. Retrieved August 05, 2019, from DARPA: https://www.darpa.mil/program/explainable-artificial-intelligence

DNV GL. (2018, September 01). *DNVGL-CG-0264 Autonomous and remotely operated ships.* Retrieved July 15, 2019, from http://rules.dnvgl.com/docs/pdf/dnvgl/cg/2018-09/dnvgl-cg-0264.pdf

EU GDPR.ORG. (2018, January 01). Retrieved August 05, 2019, from EU GDPR.ORG: https://eugdpr.org/

Gilpin, Bau, Yuan, Bajwa, Specter, & Kagal. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. *The 5th IEEE International Conference on Data Science and Advanced Analytics (DSAA 2018).* arXiv.org. Retrieved from https://arxiv.org/abs/1806.00069

Herman, B. (2017). The Promise and Peril of Human Evaluation for Model Interpretability. *NIPS 2017 Symposium on Interpretable Machine Learning.* arXiv.org. Retrieved from https://arxiv.org/abs/1711.07414

Lipton, Z. C. (2017). The Doctor Just Won't Accept That! *NIPS Symposium on Interpretable ML 2017.* arXiv.org. Retrieved from https://arxiv.org/abs/1711.08037

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems 30 (NIPS 2017).* NIPS. Retrieved from https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

NTNU. (2019). *Autoferry - Cross Disciplinary Research*. Retrieved July 24, 2019, from NTNU - Norwegian University of Science and Technology: https://www.ntnu.edu/autoferry

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 30, NO. 3, MAY 2000*, 286-297.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., . . . Wellman, M. (2019, April 1). Machine Behaviour. *Nature, 568*, 477-486.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations.*

SAE International. (2018a, June 15). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. Retrieved from https://www.sae.org/standards/content/j3016_201806/

SAE International. (2018b, December 11). *SAE International Releases Updated Visual Chart for Its “Levels of Driving Automation” Standard for Self-Driving Vehicles*. Retrieved from Society of Automobile Engineers (SAE) International: https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *2017 IEEE International Conference on Computer Vision (ICCV).*

Semcon. (2019, July 15). *WHO SEES YOU WHEN THE CAR DRIVES ITSELF?* Retrieved from Semcon web site: https://semcon.com/smilingcar/

Taddeo, M. (2017). Trusting Digital Technologies Correctly. *Minds and Machines, 27*(4), 565-568. Retrieved from https://doi.org/10.1007/s11023-017-9450-5

Wood, M., Robbel, P., Maass, M., Tebbens, R. D., Meijs, M., Harb, M., . . . Schlicht, P. (2019). *Safety First for Automated Driving.* Retrieved from https://newsroom.intel.com/wp-content/uploads/sites/11/2019/07/Intel-Safety-First-for-Automated-Driving.pdf

The Risks of Remote Pilotage in an Intelligent Fairway – preliminary considerations

Janne Lahtinen*, Osiris A. Valdez Banda , Pentti Kujala and Spyros Hirdaris
Aalto University, School of Engineering, Marine Technology Group, Espoo, Finland

ABSTRACT

To date, academic research on intelligent shipping has explored risks associated with navigation solutions embedded on-board ships. Consequently, much less research focus has been drawn on understanding challenges associated with the utilisation of new technologies in the fairway and the infrastructure surrounding a ship operating as an autonomous system of systems. Intelligent fairway is a complex emerging concept and there are no standards and/or guidelines that describe considerations on risks, emerging technology features and what facilities should offer. This paper reviews some risks of relevance to remote pilotage in Rauma 12-meter conventional fairway based on industry best practices and accident statistics. It is concluded that the transformation from conventional to remote pilotage mainly relates to challenges related to de-risking and implementing technologies, as well as the need to develop modern risk management systems and unified regulations.

Keywords: Risk analysis; intelligent fairway; situation awareness; remote pilotage operations

1 INTRODUCTION

In today's changing maritime world new new techniques change and/or improve the status quo of safe pilotage operations and surrounding fairway infrastructure. For example, developments in the accuracy of geographical positioning data, our ability to model in 3D the underwater profile of fairways, AtoN-buoys (Aids to Navigation), enhanced GPS positioning and environmental data services have opened new possibilities to surround vessels with improved safety information. Big data analytics could assist in terms of developing new generation decision support systems that help with real time management of risks in pilotage operations under remote or environmentally challenging conditions (e.g. high wind and sea state, low visibility, etc.).

However, intelligent system advances (e.g. dynamic under keel clearance management, bathymetry modelling, real time ice and sea state information, remotely controlled navigation, etc.), imply challenges resulting from their interplay with unchartered operational practices (Hollnagel, 2014). This is the reason why it is important to understand the risks associated with technologies of relevance and their impact on intelligent pilotage operations (Basnet et al., 2019).

An intelligent fairway aims to provide additional means of enhancing navigators and VTS (Vessel Traffic Service) operators' situation awareness. Yet, there is no standard that would describe, what features a fairway should offer to be justified as "*intelligent*" or "*smart*" and what this means from an assurance perspective. This paper presents a review of emerging risk management considerations based on industry practices and accident statistics of direct relevance on the Rauma intelligent fairway. Special attention is attributed to short listing the factors that may influence the rational selection of risk assessment methods and future safety management systems.

---

* Corresponding author: +358407570021, janne.p.lahtinen@aalto.fi

2 REMOTE PILOTAGE OPERATIONS

Remote pilotage places pilot ashore instead of a vessel bridge. To date, different forms of radar shore-based piloting has been conducted for decades for selected vessels in particular fairways under favourable environmental conditions. Piloting takes place in VTS operated waterways such as narrow lanes and congested waters where tolerance for error margins is minimal. In such scenarios recognising and managing human factor causation risks is critical (Grech et al.,2002 and Sandhåland, 2015). Endsley (1995) defines situation awareness as "*the perception of information elements in the surrounding, comprehension of their meaning and projection of their future status*". This is supported by the idea that remote pilotage in a standard fairway is "*an act carried out in a designated area by a pilot licensed from a position other than on-board the vessel concerned*" (Hadley, 1999). In this sense, robust situational awareness requires "*good practice*"; i.e. good communication skills and robust feel of the vessel under inadequate radar images and/or lack of vessel movement data (Lappalainen et al., 2014). To minimise risks in navigation, over the years various technologies have been implemented in IMO regulatory instruments (e.g. IMO, 1968; IMO, 2017a). Examples of technologies in use are the VHF equipment for communication, radar images, GPS-receivers combined with ECDIS (Electronic Chart Display and Information System) for chart imaging and gyrocompass for heading information, etc. Such technologies support in general the conventional chain of communication between pilot, captain and duty officer. Remote pilotage in intelligent fairway aims to enhance safety of navigation by improving the situation awareness of remote pilotage via infrastructure embedded in the intelligent fairway. Removing the pilot form the vessel and placing him in a control room ashore in itself does not improve safety. Yet, enabling a pilot's situational awareness by placing him in a working environment that offers holistic understanding of all events both on board and around the vessel, seems a tempting option to improve safety of navigation.

3 INTELLIGENT FAIRWAY

An intelligent version of a fairway could consist of "*smart*" technologies that offer better display of existing information (e.g. data from virtual base station networks) and "*intelligent*" hardware or software (e.g. real time water depth data, autonomous systems, etc.) that help to acquire new data for remote decision support. Placing the pilot ashore changes elements of communication dramatically because of risks related with loss of physical feel of the vessel, the dominance of audible and visual observations of navigation surroundings, etc. In remote conditions the need to monitor crew behaviour is compensated, as far as practical, by instrumental aids enabling broad situational awareness of the entire navigation area (Wild, 2011). As an example, Figure 1 presents the information cycle in an "*intelligent fairway*" context. It suggests that the remote pilot's decision is based on conventional instrumental tools and, in the absence of non-instrumental tools, novel data should be utilised. Accordingly, the following sections present key risk management considerations with particular focus on system complexity, the need for novel risk assessment methods and use of modern safety management systems.

3.1 Human in the loop and emerging risk methods

Risk averse remote pilotage assumes smooth operations within a complex environment ("a *socio-technical system"*) that encompasses non-linear relationships and effects between people, business processes and technologies. These interactions depend on behaviour, self-organization, robustness, emergence, hierarchical organisation, and numerosity of components. The introduction of *"intelligent"* fairway implies the possibility to reduce human errors by exposing less people to risks. Fundamental to this is that humans will remain in the loop even by distance and risks will simply migrate ashore. A practical example is given by Bruno & Lutzhoft (2010) who discuss the importance of sufficient and standardized communications between vessel crew and shore-based assistance. From a risk management perspective such effects cannot be captured adequately by the traditional risk theoretical and methodological viewpoints that are based on the principles of reductionism. Instead, emerging risk assessment methods that recognise the interplay of both organisational and environmental factors may be more appropriate. For remote pilotage this means that pilots, VTS operators and vessel crew are actors and controllers interacting within the context of an ecosystem that naturally combines both technological and business risks (Brooks et al., 2016).

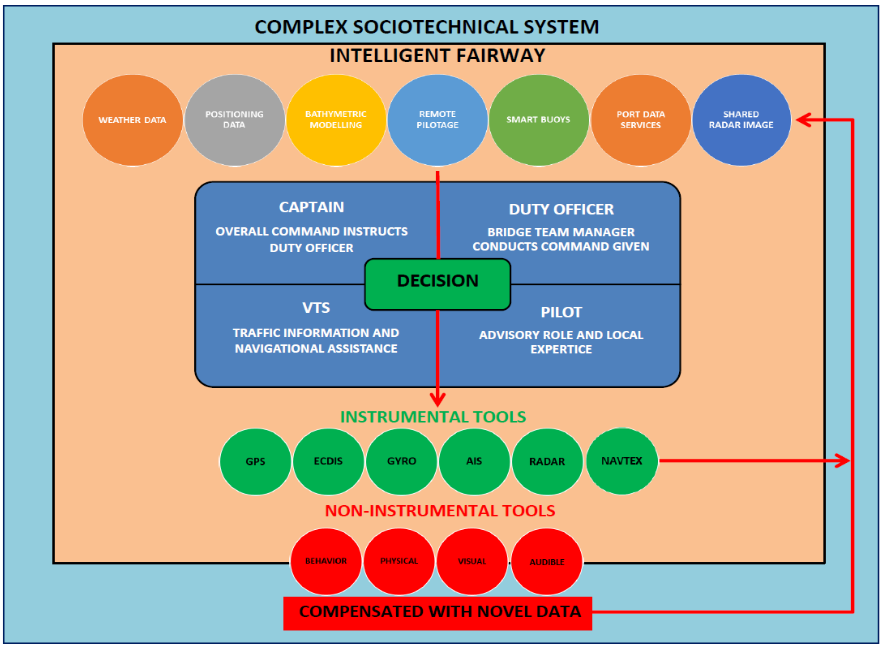


Figure 1. Communication and support to decision making for pilots.

3.2. The role of Safety Management Systems (SMS)

Pilotage organizations are not shipping companies, but still pilots must maintain compliance with various shipping regulatory elements, such as IMO regulations (e.g. IMO,1968; IMO; 2004; IMO 2017a,b,c); especially safety provisions under the International Safety Management - ISM code (IMO, 2018) and ISPS (IMO, 2003). Pilotage organizations favour the use of widely acknowledged standards of relevance (e.g. IMO,2003; ISO 9001; ISPO, 2018a,b; EMPA, 1998). This variation in terms of compliance against quality standards relates to the lack of compatibility of nationally acknowledged regulatory systems against global - mainly ISM code based - regulatory requirements (Lappalainen et al., 2014).

The ISM Code (IMO, 2018) can be understood as a form of a user manual for the vessel. It offers a variety of instructions on how to act and operate in routine and emergency situations. In turn, approved Safety Management Systems (SMS) demonstrate a practical approach against mandatory functional safety requirements on the basis of ALARP (As Low As Reasonably Practical) risk based approaches (DNVGL, 2017). Accordingly, items of non-compliance are assessed by traditional risk assessment methods (e.g. Failure Mode and Effect Analysis - FMEA and Fault Tree Analysis - FTA) and within the context of Formal Safety Assessment. However, such risk assessment methods are limited in terms of explaining the causality of accidents in modern complex systems operating in remote, autonomous or harsh environmental conditions and cannot assist with organisational risk management (Leveson, 2011). An example of this deficiency is presented by Valdez-Banda and Goerlandt (2018) for the case of winter navigation where conventional reasoning for accidents links up mostly with operator errors. Figure 2 presents a preliminary view on structural and functional requirements for future SMS. Conventional safety management system assumes that risks mostly relate with operators' actions (i.e. functional risks). Yet, if piloting is relocated ashore, risks and their mitigation will also migrate. Hence, future pilotage will still need similar policies, procedures and instructions. However, both functional and structural requirements for future pilotage should be reformatted to account for advanced use of technology. In an "*intelligent fairway*" the following issues may be considered as critical:

- *Lean reporting on risk management procedures should be faced as an emerging SMS functional requirement* that helps to stipulate and analyse transparently occurred accidents and near misses in a way that accurately measures safety. Well framed documentation protocols and reporting functions for various operational failures and succeeds embedded in traditional SMS systems are useful. However, flexibility and ability to react quickly to changes by suggesting corrective actions and safeguards is also essential under harsh environmental conditions, high automation and remoteness.

- *Optimum risk management of complexity in decision making should be faced as a SMS structural requirement and link more with practical experience.* However, in light of the impact of emerging technologies in complex sociotechnical ecosystems further research on the adaptation of novel risk assessment methods with the aim to identify gaps in system performance under environmentally demanding or remote conditions may be useful.
- *Remoteness or automation in operations should link up SMS structural requirements on human resource management and acknowledge technology know how as a functional requirement.* This is because the use of novel technology does not necessarily reduce or eliminate the need to invest in human capital. To maintain know-how on system operations, training and retaining personnel should be carefully considered and adapted to emerging needs.
- *Lean procedures and audits should be embedded as SMS structural requirements with the aim to manage compliance against legislation, regulations and example terms of insurance.* Standard organization management practices tend to use auditing to ensure that risks are assessed in the way intended when the system was first established. Complexity of remote pilotage operations imply that audits should be maintained satisfying internal quality control needs and external requirements (e.g. classification societies, port state control, insurance) as well as client expectations. As part of this process future risk assessment systems should be practical and recognize data and information sources, and exchanges between parties. This relates to both ease of use and cost efficiency and should be accounted for in employee performance monitoring standards (e.g. Key Performance Indicators - KPI).
- *Future SMS systems should functionally encourage the transparent communication of lessons learnt and progress.* To achieve this, new methods should be validated by intelligent interpretation of statistics of accidents (or in general unwanted events), the impact of improved working conditions on and efficiency at work (KPI). The role of STAMP risk assessment methods in this respect could be beneficial.

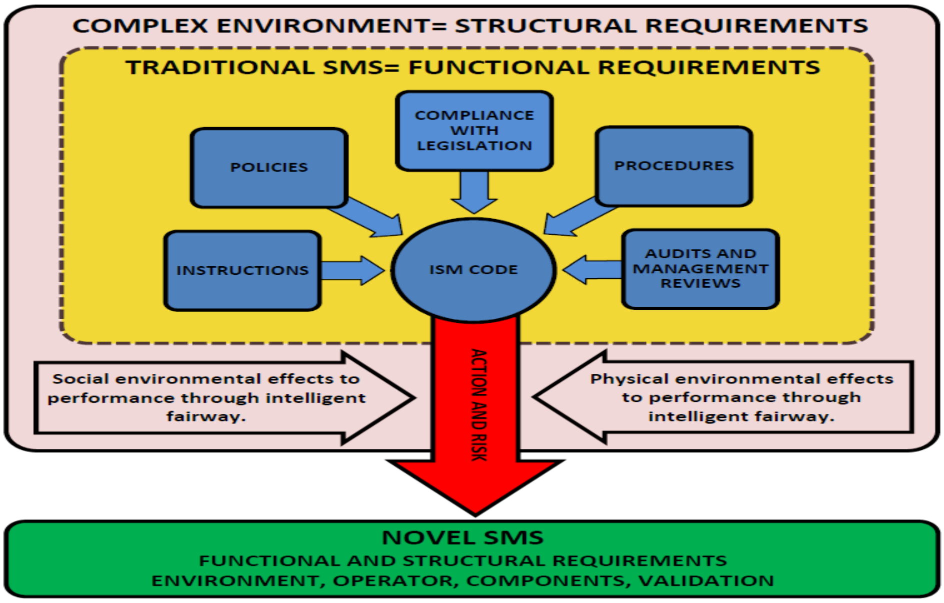

COMPLEX ENVIRONMENT= STRUCTURAL REQUIREMENTS
TRADITIONAL SMS= FUNCTIONAL REQUIREMENTS
COMPLIANCE WITH LEGISLATION
POLICIES
PROCEDURES
INSTRUCTIONS
ISM CODE
AUDITS AND MANAGEMENT REVIEWS
ACTION AND RISK
Social environmental effects to performance through intelligent fairway.
Physical environmental effects to performance through intelligent fairway.
NOVEL SMS
FUNCTIONAL AND STRUCTURAL REQUIREMENTS
ENVIRONMENT, OPERATOR, COMPONENTS, VALIDATION

Figure 2. Components of the ISM code (blue boxes) encompass traditional features of SMS (yellow box). Actions and their associated risks (red) are realized in a complex environment (purple box) and may be subject to social and environmental factors. Thus, a novel SMS should consider both functional and structural requirements (green box).

4 THE CASE OF RAUMA 12 METER FAIRWAY

Motivated by the principles outlined in Section 2 we hereby discuss unknown risks related to remote pilotage for a practical case of relevance; namely the Rauma 12 meter fairway (Figure 3). Studies on this fairway are carried out as part of ISTLAB (Intelligent Shipping Technology Test Laboratory) project led by the Satakunta University of Applied Sciences of Finland (SAMK) and funded by the EU Regional Development Fund (https://www.maanmittauslaitos.fi/en/node/12089). The goal of the project is to establish an open innovation laboratory for the development of a smart navigation fairway at the Finnish port of Rauma. It therefore aims to merge features of the navigation simulator of SAMK with the Finnish Transport Agency's bathymetric model, smart buoy and sea current monitoring, the Finnish Geospatial Research Institute's navigation system (FGRI, 2019) and the Finnish Meteorological Institute's survey of wave and ice conditions (SAMK, 2019). The statistics presented are based on incidents and near misses recorded by West Coast VTS and Traficom and refer to vessels under pilotage by a professional pilot or a captain with pilotage exemption certified by Traficom. It is worthwhile noting that vessels with pilotage exemption are regular visitors with crews highly experienced in manoeuvres under various environmental circumstances and traffic scenarios. The case of a collision resulting from poor situational awareness is especially highlighted. This is because statistical data review demonstrated that this is a key route cause for accidents.

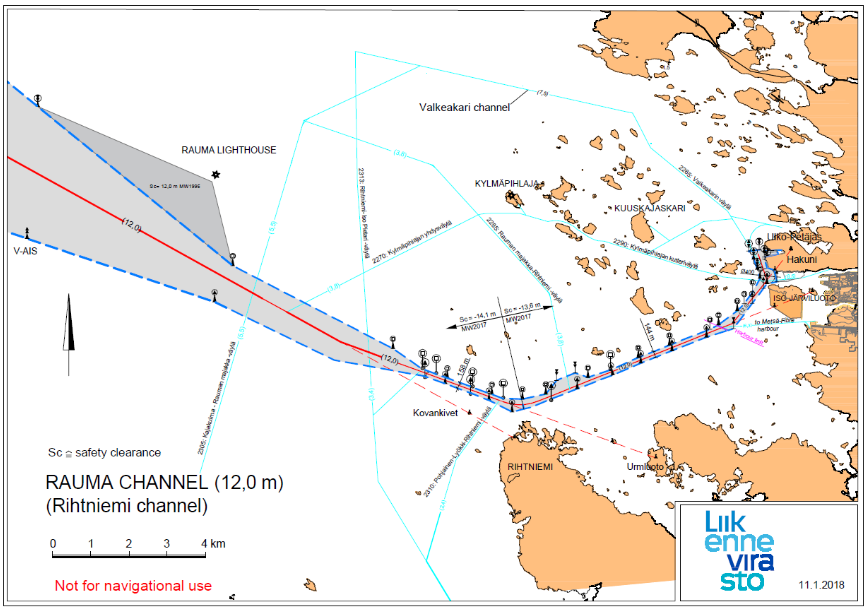


Figure 3. The Rauma 12 meter fairway, (FTIA, 2019).

4.1 Review of Accident statistics

Between 2014 and 2019 the Port of Rauma 12m conventional fairway was visited by 5,592 vessels (Port of Rauma, 2019). Figure 4 summarises the reported non-conformities to West Coast VTS that are linked to the detected navigational risks in the conventional Rauma fairway. Those are classified as shallow water, technical failure, communication, collision danger and exit from the fairway area. Whereas in all of these cases clear instructions must be given to manage operational risks, shallow water proximity poses the worst-case scenario. Research indicated that cases where VTS signalled a warning of shallow waters also link with poor communication. Currently, VTS has no standard on level of decision-making support for vessel crews or alternative means to report dangerous situations. Decisions depend on the individual VTS operator. Instead, navigation choices are at the discretion of the navigators (certified captain or pilot) who respond on the basis of experience and report on events that in their view could develop into severe accidents. Also, periodical reporting is conducted in a format that is not regular in terms of duration or time of the year and therefore it cannot provide any consistent traces.

Between 2009 and 2019 11,935 vessels visited the Port of Rauma (Port of Rauma, 2019). Review of Traficom statistics for this period (Figure 5) indicates that grounding, collision and capsizing have been the main accident variants. It is worthwhile noting that of the five cases of grounding, four occurred with a pilot on board and one when the certified vessel master was in charge. The piloted vessels experienced two groundings due to human errors and further two as a result of technical failure. The pilot exempt vessels experienced one case of grounding that resulted from human error. The three recorded collisions occurred with a certified master on-board. Of those two occurred under bad environmental conditions and one was attributed to human error. The capsized vessel accident refers to the case of a dredging work vessel on a crew transport operation. It is worthwhile to note that this was not a vessel of cargo specification and was not under pilotage. Of the five occurred groundings three were due to human error and two could be attributed to technical failure. This review indicates that the presence of a pilot does not necessarily imply automated ticket to safe navigation or the contribution of a knowledgeable bridge team familiar with localities and their own vessel. Yet, the majority of the vessels that operate in this area are piloted safely by professional pilots meaning that the prevailing safety management process generally functions well.

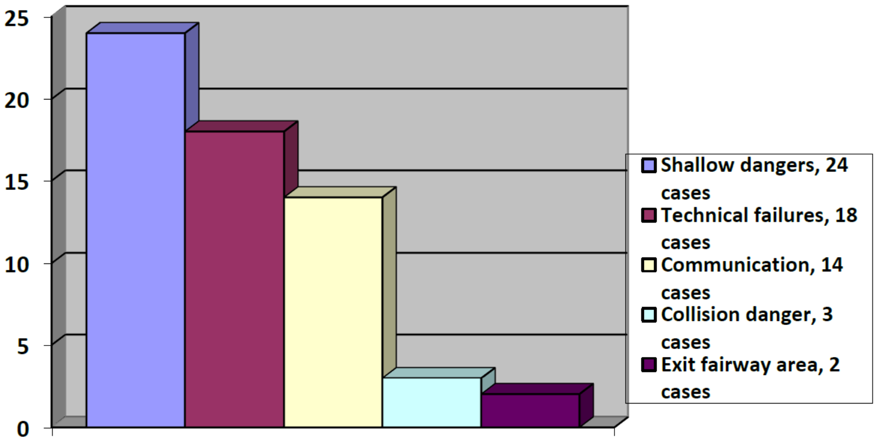


Figure 4. Non-conformities in Rauma fairways 2014-2019 reported by West Coast VTS.

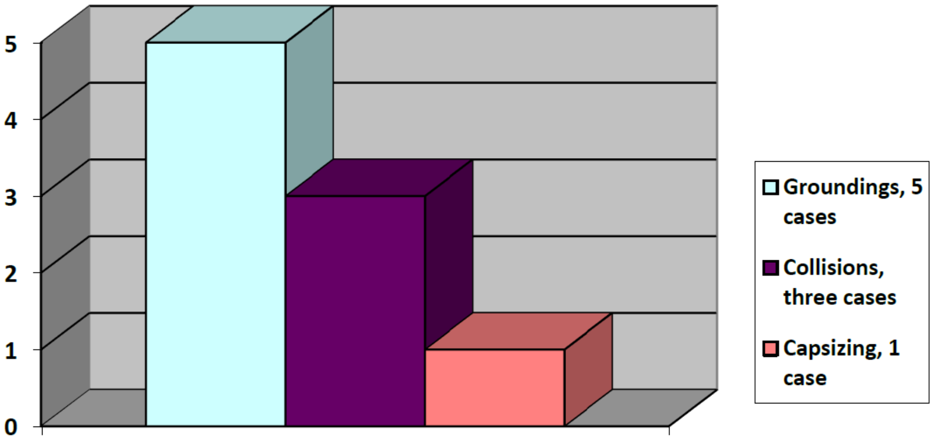


Figure 5. Accidents in Rauma fairways 2009-2019 reported by Traficom.

4.2 Situational awareness - a key risk

The collision between the Norwegian frigate “*KNM Helge Ingstad*” and the oil tanker Sola TS in Norway on 8.11.2018 at 04:01 local time demonstrates that human presence on site with insufficient understanding of the events may significantly affect decision making. The accident was a complex chain of events that involved numerous factors, such as bridge crews, pilot, VTS and their interaction (NAIB, 2018). The case of KNM Helge Ingstad suggets that situation awareness can be better ashore that on board, provided that there is sufficient data availbale to support decision making.

At the time of the accident, "*KNM Helge Ingstad*" was sailing southbound toward the Sture oil terminal on their starboard side. Her AIS was set in receiving mode, meaning other vessels or Fedje VTS could not access her name or call sign. At the same time "*Sola TS*", with a pilot on board, departed from Sture terminal. Shortly after departure the pilot noticed "*KNM Helge Ingstad*" on radar visibly showing her green light. The pilot assumed that this other vessel was on her way cross over "*Sola TS*" navigation line and asked for her identity from Fedje VTS who at first were not able to identify her. With some brief delay they informed the pilot of "*Sola TS*" that the other vessel could be "*KNM Helge Ingstad*". For this reason, "*Sola TS*" made contact with "*KNM Helge Ingstad*" requesting them to change their course to starboard. The officer of the watch on "*KNM Helge Ingstad*" declined this action assuming that an object on her starboard side prohibits course change. Despite further calls the frigate maintained her course and carried out a collision avoidance maneuver only the last moment. This resulted in collision between the two vessels.

The accident investigators did not discover any technical equipment malfunction as a reason (NAIB, 2018). It was concluded that at the time of the accident "*Sola TS*" was seen from "*KNM Helge Ingstad*" bridge against terminal lights. For this reason, the frigate failed to recognize her and instead assumed that a tanker vessel obstructed her to conduct a starboard maneuver. Fedje VTS, pilot on board "*Sola TS*" as well as the bridge crew on board "*KNM Helge Ingstad*" presumed the accident each in their own way. However, "*KNM Helge Ingstad*" bridge crew with obligation to take evasive action was the only party that did not comprehend the developing situation.

The accident demonstrated that from a situational awareness perspective there was a gap between perception and comprehension. Navigational scenario, such as the case of KNM Helge Ingstad and Sola TS being in crossing trajectories, is an event that includes various parties making decisions based on justified assumption that other affected party will act in a certain manner to clear the situation safely. Safe navigating can then be seen as a team decision making, although traditionally we see the navigation team consisting of navigators on a bridge of a single vessel. Poor situational awareness of one team member invalidates decisions made by the remaining team (Endsley, 1995). Bruno and Lutzhoft (2010) conducted interviews of VTS operators in their research for human aspects in shore-based assistance for ships concluding importance of communication and trust between shore operator and vessel crew. Their interview results and case of ''*KNM Helge Ingstad*'' support the statement that placing pilot, or some level of control of vessel to the shore with holistic situation overview may increase the level of safety by reducing risks associated with human factors.

5 DISCUSSION

The accident statistics discussed in section 3 suggest that: (a) although piloting has the potential to free vessels from risks associated to unwanted events local knowledge does not free pilot from the weaknesses of human ability to percept and comprehend his surroundings. However, pilots may lose situation awareness while certified masters could have reasonable potential to deliver upon expectations; (b) the frequency of VTS interference to vessel navigation in relation to occurred accidents in Rauma fairway demonstrates strong relationship between situational awareness of vessel command with communication between different parties involved (vessel crew, pilot, VTS). For these reasons, piloting vessels in intelligent fairways should become an operation that recognises VTS as part of the risk management process and in support of these future marine operations will have to account for intelligence in a form of novel technology for traffic management. Improved SMS functionality potentially helps with improved communications between VTS and the vessels (see Figure 2).

With particular reference to technology development and implementation, suitable use of all stakeholders involved and their interaction is the key. New technologies (e.g. big data analytics) if tested and assured could offer navigators improved understanding of their surroundings. This is the reason why better lines of quality communication must be established between classification societies, ports, regulators and operators including pilotage organisations. For example, real time water depth information may allow vessels to enter and leave port with widened time window for under keel clearance making both the port and the fairway use considerably more efficient and safer. This complementary feature is highlighted in ports with tidal range. However, implementation of such information for practical decision support under ECDIS platform poses a number of risks in terms of technology development, testing, implementation and approval. On the technology front an ECDIS platform with enhanced real time water depth information could be the obvious solution. Accordingly, real time information of water depth readings that present water levels above the chart datum would have to be implemented to allow navigators to make decisions of under keel clearances that are usually stated as : *"the minimum water depth under the keel in SMS"*. In turn, bathymetry modelling of the fairway combined with real time water depth data could give exact presentation of water levels within fairway region. Finally, testing, implementation and approval would require training of and extensive collaborative work between end user groups, broad involvement of engineers and social scientists that have to mitigate risks associated with challenges on human behaviour and ergonomics, class societies to certify decision support systems and regulators who would raise the standard by demanding implementation for practical use.

Practically there may be some possibility that failures will occur. Yet, their effects should not compromise system integrity and operation. For this reason, future risk management methods should reveal the causality hidden in the relationships between different participants as system components. Further testing and validation of STAMP (Systems Theoretic Accident Models & Processes) could help define efficient safety control options that in turn increase system resilience.

6 CONCLUSIONS

An intelligent fairway aims to provide additional means of enhancing navigators and VTS operators' situational awareness. This paper described some of the features it could offer to be classified as "intelligent" by reviewing risk management practices and accident statistics in way of the Finnish Port of Rauma. It was concluded that the transformation from conventional to remote pilotage mainly relates to challenges related to technology as well as the lack of modern risk management systems and unified regulations. Future developments should therefore focus on of the implementation of modern risk assessment methods for the validation of emerging technologies and the development of an SMS able to accommodate for all autonomous operations in remote conditions. With reference to the former it is suggested that subject to further testing and validation STAMP methods that focus on safety constraints while still acknowledging failures of individual components could be the best way forward in terms of managing dynamic risks associated with technology development, validation and implementation. Future SMS should consider both structural and functional requirements and make suitable use of all stakeholders involved and their interactions.

ACKNOWLEDGEMENTS

The research presented in this paper is sponsored by Satakunta University of Applied Sciences and Aalto University, Finland and Intelligent Shipping Technology Test Laboratory project (https://www.maanmittauslaitos.fi/en/node/12089). The authors would like to express their gratitude to Sari Talja and Kati Westerlund of VTS Finland and Esa Pasanen of Traficom.

REFERENCES

Basnet, S., Valdez Banda, O. and Hirdaris, S. (2019). The Management of Risk in Autonomous Marine Ecosystems – Preliminary Ideas. In MA Ramos, C Thieme, IB Utne & A Mosleh (Eds.), Proceedings of the First International Workshop on Autonomous Systems Safety (IWASS'19). Norwegian University of Science and Technology, pp. 112-121, Trondheim, Norway.

Brooks, B., Coltman, T., and Miles, Y. (2016). Technological Innovation in the Maritime Industry: The Case of Remote Pilotage and Navigational Assistance. *Journal of Navigation,* 69:777-793 .

Bruno, K., and Lutzhoft, M. (2010). Virtually Being There: Human Aspects of Shore-based Ship Assistance. *WMU Journal of Maritime Affairs,* 9(1):81–92

DNVGL. (2017). *Risk Management in Marine and Subsea Operations.* Recommended practice DNVGL-RP-N101.

EMPA (1998). Code of Best Practice for European Maritime Pilots, 33rd EMPA General Meeting, European Maritime Pilots Association.

Endsley, M. R. (1995). Measurement of Situation Awareness in Dynamic Systems. *Human Factors - The Journal of Human Factors and Ergonomics Society,* 37(1):65-84.

FGRI (2019). *Finnish Geospatial Research Institute review of ISTLAB* (Information retrieved from https://www.maanmittauslaitos.fi/en/node/12089).

FTIA (2019). The *Rauma 12m Channel Fairway Card* (Information retrieved from https://vayla.fi/documents/21386/135676/Rauman+12+m+v%C3%A4yl%C3%A4+eng.pdf/fa9033f5-7d89-4d61-bf0a-aaa8384cbccd).

Grech, M. R., Horberry, T., and Smith, A. (2002). Human Error in Maritime Operations: Analyses of Accident Reports Using the Leximancer Tool. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(19):1718–1721.

Hadley, M. (1999). Issues in Remote Pilotage. *Journal of Navigation, 52*(1):1-10.

Hollnagel, E. (2014). *Safety-I and Safety-II. The past and Future of Safety Management* Surrey, England: Routledge, 1st Edition ISBN-13: 978-1472423085.

IMO (1968). *Recommendation on port advisory services.* Resolution A.158 (ES.IV), The International Maritime Organization, London, UK.

IMO (2003). *The International Ship and Port Facility Security Code*. The International Maritime Organization, London, UK, ISBN 978-92-801-5149-7.

IMO (2004). *Recommendations on training certification and on operational procedures for maritime pilots other than deep sea pilots*. Resolution A23/Res.960 The International Maritime Organization, London, UK

IMO (2017a). *Safety Of Life At Sea*. The International Maritime Organization, London, UK, ISBN 978-92-801-1594-9.

IMO (2017b). *International Convention for the Prevention of Pollution from Ships*. The International Maritime Organization, London, UK, ISBN 978-92-801-1657-1.

IMO (2017c). *International Convention on Standards of Training, Certification and Watchkeeping for Seafarers*. The International Maritime Organization, London, UK ISBN 978-92-801-1635-9.

IMO (2018). *International Safety Management Code* London: The International Maritime Organization, London, UK, ISBN: 978-92-801-1696-0.

ISO (9001). *International Organisation for Standardisation, ISO 9001:2015(en) Quality Management Systems – Requirements.*

ISPO (2018a). *Guidelines and Additional Information to the ISPO International Users Group* (Information retrieved from https://www.ispo-standard.com/Downloads.aspx).

ISPO (2018b). *International Standard for maritime Pilot Organizations International Users Group* (Information retrieved from https://www.ispo-standard.com/Downloads.aspx).

Lappalainen, J., Kunnaala, V., and Tapaninen, U. (2014). Present Pilotage Practices in Finland. *WMU Journal of Maritime Affairs*, Vol.13(1), 1-23.

Leveson, N. (2011). *Engineering a Safer World - Systems Thinking Applied to Safety.* Cambridge: MIT Press, ISBN: 9780262016629.

NAIB (2018). *Preliminary Marine Accident Investigation Report on the Collision between the Frigate "KNM Helge Ingstad" and the Oil Tanker "Sola TS".* Lillestrøm: Accident Investigation Board Norway (Information retrieved from https://www.aibn.no/Marine/Investigations/18-968?iid=25573&pid=SHT-Report-Attachments.Native-InnerFile-File&attach=1).

Port of Rauma (2019). *Port of Rauma Vessel Traffic Statistics* (Information retrieved from: https://www.portofrauma.com/sites/default/files/rauman_satama_liikennetilastot_1991-2018.pdf).

Sandhåland, H., Oltedal, H. A., Hystad, S. W. and Eid, J. (2015). Distributed Situation Awareness in Complex Collaborative Systems: A field Study of Bridge Operations on Platform Supply Vessels. *Journal of Occupational and Organizational Psychology*, 88:273-294.

SUAS (2019). *ISTLAB – Intelligent Shipping Technology Test Laboratory* (Information retrieved from https://www.samk.fi/tyoelama-ja-tutkimus/hankkeet/).

Valdez Banda, O. A., and Goerlandt, F. (2018). A STAMP based approach for desingning maritime safety management systems. *Safety Science, 109*:109-129.

Wild, R. J. (2011). The Paradigm and the Paradox of Perfect Pilotage. *The Journal of Navigation, 64*(1), 183-191.

A Targets Detection Approach Based on an improved R-CNN Algorithm for Inland River Crossing Area Marine Radar Image

Chao Wu[1,2,4,*], Qing Wu[1,2] and Shuwu Wang[2,3]
[1] School of Logistics Engineering, Wuhan University of Technology, Wuhan, Hubei
[2] National Engineering Research Centre for Water Transport Safety, Wuhan, Hubei
[3] School of Energy and Power Engineering, Wuhan University of Technology, Wuhan, Hubei
[4] School of Electronic and Information, Yangtze University, Jingzhou, Hubei

ABSTRACT

The vessels sailing in the inland river are potential fatal threat to the ferries. In order to ensure the navigation safety, ferries must be able to effectively perceive other dynamic targets in real time with the help of navigational AIDS. In this study, an improved R-CNN algorithm was proposed to detect the targets in radar images. At the very beginning, positive and negative sample sets have been created by manual from a large number of radar images. Subsequently, the CNN network is used to extract the features from the positive and negative sample sets. Taking the updated radar image as input, the radar image pre-processing is completed, and then the region proposal is obtained by calculating the connected region in the radar image based on breadth-first search algorithm. After adjusting the fixed size of the region, the image is sent to the network which has been trained and tested to reach a stable state to extract the features. The classification of target is determined by SVM, and the region proposal position is refined adjusted by regression and the result is output. The proposed approach is unique in that the maritime radar image is used as the research object, which is different from the video image or optical image used in other studies. Particularly, the validity and accuracy of the proposed approach are verified by testing the data collected in the field.

Keywords: Marine Radar Image, Target Detection, R-CNN, Navigation Safety

1 INTRODUCTION AND BACKGROUND

Ferries are one of the means of transportation in the Yangtze River basin of China. Therefore, ferry has long been the key supervision object of water traffic safety management department. Regrettably, in numerous maritime accidents, inland river ferries account for a larger proportion. The main reason for ferry accidents is that ferries have to work across the entire inland waterway, where there are often a large number of vessels sailing. In order to ensure the navigation safety of ferries and avoid disasters, ferry pilots need to use navigational AIDS to perceive the navigation environment in the transit area and detect potential threats in advance. Among many AIDS to navigation, marine radar is an indispensable one. Within the detection range, the detected target exists in the sequentially updating radar images in the form of light spot. Experienced radar operators can easily identify all kinds of targets in the waterway according to the size and moving state of the spot in the sequentially updating radar images.

With the rapid development of artificial intelligence, more and more human tasks will be replaced by computers. Compared with manual monitoring, computer has obvious advantages in monitoring navigation situation in the waterway. Using target detection algorithm to identify the key targets in radar images will save cost and improve work efficiency to a great extent. The task of target detection is to find out all the interested targets in the image and determine their positions and

---

[*] Chao Wu: chaocruise@whut.edu.cn

sizes. However, target detection has always been the most challenging problem in the field of machine vision due to the appearance, shape and attitude of various objects, as well as the interference of illumination, shielding and other factors during imaging. In recent years, as one of the three tasks in the field of computer vision, target detection has made remarkable achievements with the rapid development of computer vision technology and the promotion of deep learning. Since Girshick [1] proposed R-CNN algorithm in 2013, it has made a breakthrough in the field of target detection。 Since then, in terms of target detection, SPP Net [2], Fast R-CNN [3], Faster R-CNN [4], YOLO [5], SSD [6] have emerged successively. In practical applications, these target detection algorithms based on deep learning and constantly improved are mainly aimed at optical images or video images, and achieve good results. At present, these target detection algorithms rarely involve the processing of radar images which are obviously different from optical images.

This research proposes an approach for detecting vessels in inland waterways. This approach is for the continuous output image of marine radar, and it is improved based on R-CNN algorithm. The remainder of this article is organized as follows. In section 'Literature review', the research on target detection algorithm based on deep learning is introduced. Section 'The proposed approach' presents an improved R-CNN target detection algorithm for radar images. In section 'Case study', a case study is conducted to demonstrate and validate the proposed methodology. The conclusions of this study are presented in section 'Conclusion'.

2 LITERATURE REVIEW

Target detection is widely used in intelligent monitoring, artificial intelligence, unmanned driving, image understanding and other research fields. Most of the early target detection algorithms are based on manual features. Due to the lack of effective image feature expression methods before deep learning was proposed, people need to involve more diversified inspection algorithms to make up for the defects in manual feature expression ability. In the early target detection algorithm, the representative algorithms are viola-jones detector (VJ Detector) [7] and HOG [8] pedestrian detector and Deformable Part based model [9]. The face detection algorithm proposed by Pual Viola and Michael Jones in the article published on CVPR is VJ Detector which realized the real-time face detection for the first time in the extremely limited computer resources in 2001. The computational speed of VJ detector is dozens or even hundreds of times that of other detection algorithms in the same period, which greatly promotes the commercialization of face detection application. HOG feature is proposed to solve the problem of pedestrian detection. HOG feature is an important improvement on the histogram feature of gradient direction, and it is the basis of all target detectors based on gradient feature. HOG detector uses the original idea of multi-scale pyramid plus sliding window in operation. To detect different sizes of targets, the size of detector window is usually fixed and the image is scaled step by step to build a multi-scale pyramid. In order to take into account the operation speed and performance, the classifier used by the HOG detector is usually a linear classifier or a cascade decision classifier. Deformable Part based Model（DPM）is the culmination of the development of inspection algorithms based on classical manual features. The main idea of DPM is to split and transform the detection problem of the whole target in the traditional target detection algorithm into the detection problem of each Part of the Model, and then aggregate the detection results of each Part to obtain the final inspection results.

When convolutional neural network achieved great success in ImageNet classification task in 2012, Girshick et al. took the lead in proposing the target detection framework of regional convolutional network in 2014. Since then, the field of target detection has entered a stage of rapid development. With the deepening of convolutional neural network layers, the abstract ability, anti-translation ability and anti-scale change ability of the network become stronger and stronger. However, the key to apply convolution network to target detection effectively lies in how to effectively solve the contradiction between translation and scale invariance of deep network and translation and scale covariant requirements in target detection. In order to solve this contradiction, researchers abandoned the detection scheme based on feature graph and sliding window, and turned their attention to the algorithm of Object Proposal Detection, which is more accurate in positioning.

At present, the relevant research focuses on pedestrian detection, vehicle detection, medical image detection and other aspects, and has achieved good results. Pedestrian detection continues to hold a significant role in the concept, analysis and function of computer vision. Katleho et al. [10] evaluated a powerful deep learning technology of R-CNN based on two different pedestrian detection data sets. Deep learning feature extraction model and R-CNN detector were used in their research. The deep learning feature extraction used is the Alexnet. Transfer learning is performed on the feature extraction model to adjust the weights of the convolutional neural networks to favour classification on the selected datasets. The R-CNN detector is then trained on the deep learning feature extraction model for pedestrian detection. Different types of vehicles, such as buses and cars, can be quite different in shapes and details. This makes it more difficult to try to learn a single feature vector that can detect all types of vehicles using a single object class. Sitapa et al. [11] proposed an approach to perform vehicle detection with Sub-Classes categories learning using R-CNN in order to improve the performance of vehicle detection. Instead of using a single object class, which is a vehicle in this experiment, to train on the R-CNN, they used multiple sub-classes of vehicles so that the network can better learn the features of each individual type. Generally, R-CNN algorithm mainly detects two-dimensional images, but most medical images are three-dimensional, which increases the difficulty of target detection. Yun Chen et al. [12] present a unified framework called Volume R-CNN for object detection in volumetric data. Volume R-CNN is an end-to-end method that could perform region proposal, classification and instance segmentation all in one model, which dramatically reduces computational overhead and parameter numbers. These tasks are joined using a key component named RoIAlign3D that extracts features of RoIs smoothly and works superiorly well for small objects in the 3D image.

Although radar image is a kind of image, the researchers who can get access to real-time radar image and study it are limited to some related majors. However, radar images usually reflect the real-time situation of a certain area, so it is particularly important for supervisors to accurately identify targets in radar images. In the research of target detection in maritime radar images, Ma et al. [13] proposes a Bayesian Network-based methodology to extract moving vessels from a plethora of blips captured in frame-by-frame radar images. First, the inter-frame differences or graph characteristics of blips, such as velocity, direction, and shape, are quantified and selected as nodes to construct a Directed Acyclic Graph (DAG), which is used for reasoning the probability of a blip being a moving vessel. Particularly, an unequal-distance discretisation method is proposed to reduce the intervals of a blip's characteristics for avoiding the combinatorial explosion problem. Then, the undetermined DAG structure and parameters are learned from manually verified data samples. Finally, based on the probabilities reasoned by the DAG, judgments on blips being moving vessels are determined by an appropriate threshold on a Receiver Operating Characteristic (ROC) curve.

3 THE PROPOSED APPROACH

This research proposed a target detection approach based on an improved R-CNN algorithm for Inland River crossing area radar image. According to the principle of R-CNN algorithm, a model for dichotomy needs to be trained first. The sample sets used for the training of dichotomy model is created by artificial experience and divided into positive sample set and negative sample set. When the model reaches the stable state after training and testing, the constantly updated radar images are used as input. The constantly updated radar images need to be pre-processed first. Then, proposal regions are obtained based on the breadth-first search algorithm, which is the improvement of R-CNN algorithm in this study. After adjusting the fixed size of all the proposal regions, the trained CNN network is used to extract features from them, and the categories in the proposal regions are judged according to SVM. Finally, regression is used to modify the region proposal position and output the result. The flowchart of this approach is illustrated in Figure 1. Particularly, only real-time radar images are used as the data source.

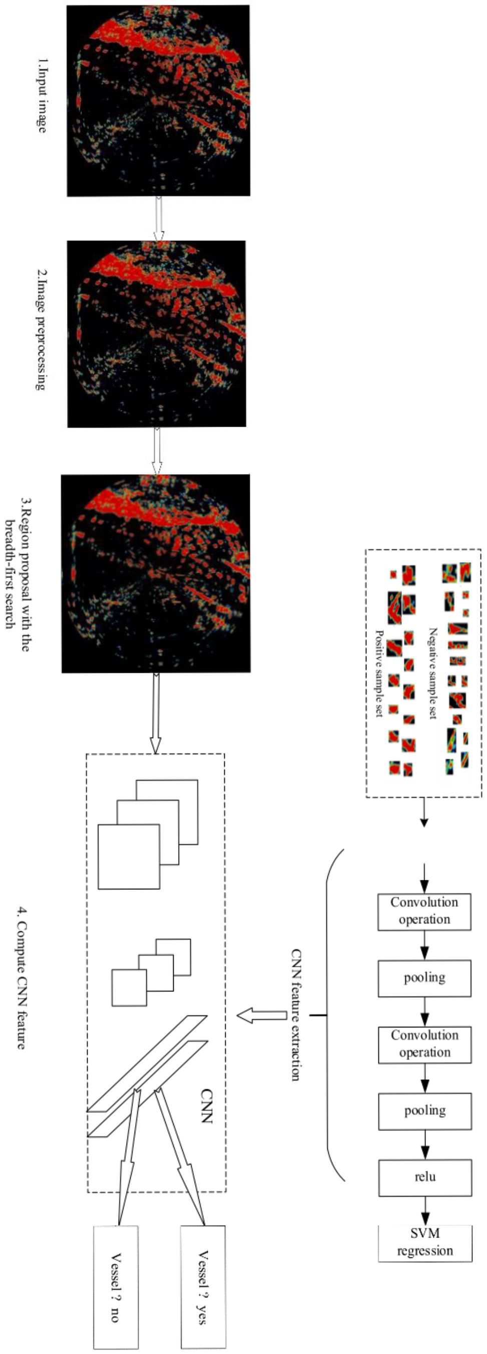


Figure 1: The flowchart of the proposed approach

3.1 Feature extraction and SVM training

For general optical images, the complexity of the target to be detected in the image is high in contour, color, texture, size, spatial overlap and other aspects. Therefore, in order to extract target feature better, the convolutional neural network structure is usually designed to be relatively complex. There are eight weighted layers in the Alexnet [14]. The structure of Alexnet is shown in Figure 2. The first five layers are the convolution layer, and the remaining three are the full connection layer. The output of the last full connectivity layer is the input of 1000 dimensional softmax. Softmax generates a distribution of 1000 categories. In addition, the detection of targets in optical images may involve multiple target categories, such as identifying people, cars, plants and animals etc. Therefore, the classification number will be set according to the actual situation in the classification model of training or downloading. In the R-CNN algorithm, the number of classifications after fine-tuning was 21. There are 20 categories and 1 background.

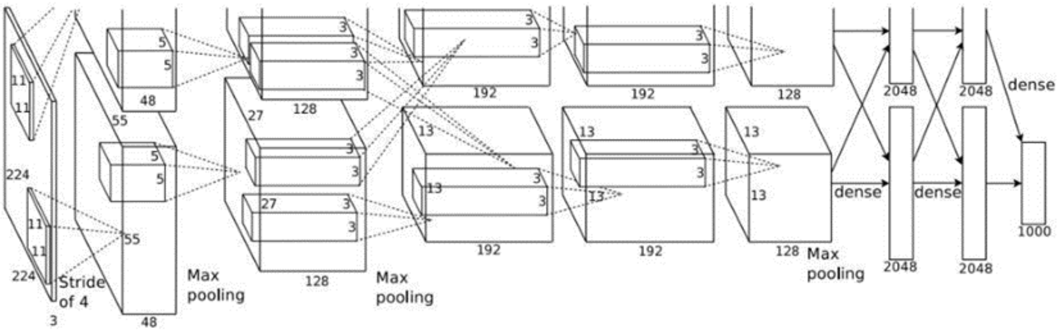


Figure 2: The structure of AlexNet

However, for the marine radar images, the targets to be detected are the vessels in crossing area. The representation of radar image is divided into foreground and background. The spot formed by the radar reflection wave in the image is considered as the suspected target or foreground, and the empty area in the image is considered as the background. Compared with the optical images, radar images differ greatly both in imaging form and in the complexity of the contour details. The size of the detected target spot in the radar image is related to the actual size of the target. The contour of the radar spot is irregular, and there are differences in the contour of the same target in the continuous output radar image. But, the vessel spot in radar image is basically fusiform and has a certain aspect ratio, while the other object spot is totally irregular. Therefore, with the support of enough positive and negative sample sets, the structure of convolutional neural network can be simplified for feature extraction of vessel spot. In this research, the convolutional neural network structure is simplified. The feature extraction network consists of two convolution layers, two pooling layers and one relu layer. The process of feature extraction and SVM training is shown in Figure 3.

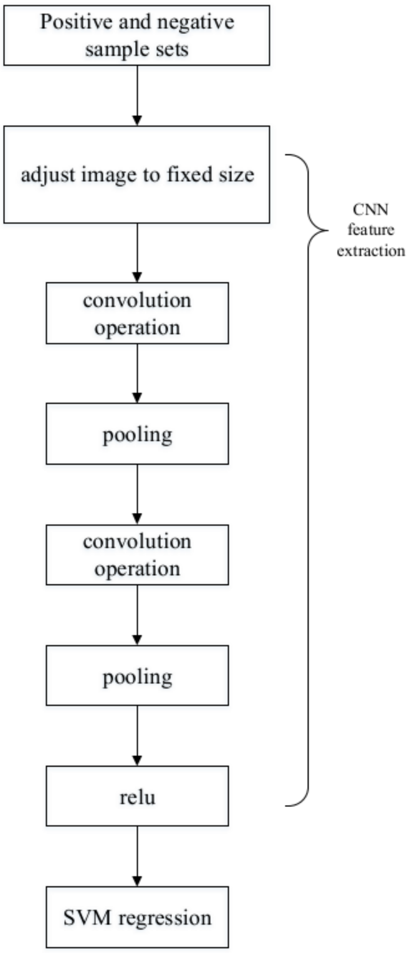


Figure 3: CNN feature extraction

As a kind of dichotomy classifier, the main idea of SVM is to find a hyperplane in the space that can divide all data samples, and make the distance between all data from samples of different categories to this hyperplane be the shortest. The principle is shown in Figure 4.

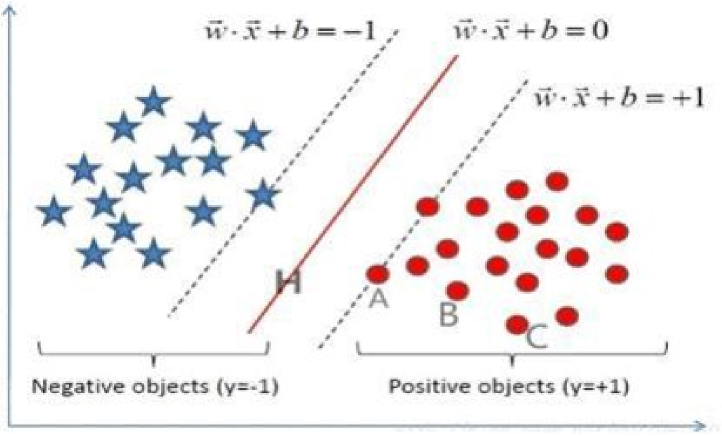

$\vec{w}\cdot\vec{x}+b=-1$
$\vec{w}\cdot\vec{x}+b=0$
$\vec{w}\cdot\vec{x}+b=+1$
H
A
B
C
Negative objects (y=-1)
Positive objects (y=+1)

Figure 4: Schematic diagram of SVM

The hyperplane is shown in equation (1)

$$\mathrm{w}^{T}x+b=0 \tag{1}$$

Suppose P(x_1, x_2, … , x_n) is a point in the sample, where x represents the ith characteristic variable, then the distance d from this point to the hyperplane can be calculated by equation (2).

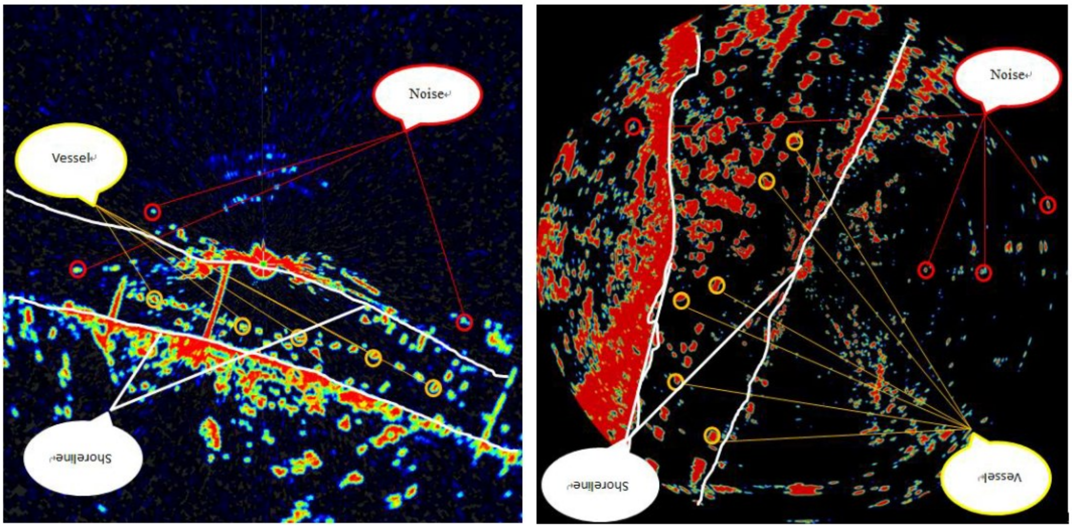
$$d = \frac{|\mathrm{w}_1 * x_1 + \mathrm{w}_2 * x_2 + \cdots + \mathrm{w}_n * x_n + \mathrm{b}|}{\sqrt{\mathrm{w}_1^2 + \mathrm{w}_2^2 + \cdots + \mathrm{w}_\mathrm{n}^2}} = \frac{|\mathrm{W}^T * X + \mathrm{b}|}{||\mathrm{W}||} \tag{2}$$

Where ǁWǁ is the norm of the hyperplane, and the constant b is similar to the intercept of the linear equation.

3.2 Radar image preprocessing

Marine radar has been widely used in all types of ships or coastal surveillance as important navigational aids. In general, the detection range of marine radar can be altered manually, so the range of marine radar detection area will be different according to the actual situation. The output images of different types of shore-based monitoring radars are shown in Figure 5 (a) and (b). It can be seen from the figures that the center of the image is the location of the radar. In a radar image, besides various targets, there are also a lot of noises. Some noises have been marked in figures. In addition, the channel has been marked. However, the channel portion of the radar image is the exactly focus area for regulators. When carrying out target detection on radar image, a large amount of additional calculation data will be generated in most areas of the image except the channel, which will seriously affect the calculation speed. Therefore, in order to minimize the computation, the original radar image must be pre-processed.

The general image pretreatment method includes binarization, dilation and erosion etc. But these pretreatment methods are not suitable. The reason is that although these operations can reduce the noise in the image to a certain extent, the noise points in the radar image are relatively small, which will not affect the detection of the target, or even can be ignored. In addition, dilation and erosion operations will make the contour of the spot in the radar image become smooth, which is not conducive to feature extraction. Therefore, in this study, the pretreatment of radar image is only to use the mask to filter out the channel, which only has the signal in the channel. This reduces the amount of data and improves the computing speed.



(a) (b)

Figure 5: The output images of shore-based monitoring radar

3.3 Region proposal

Traditional target detection algorithms have low efficiency. The inefficiency is mainly caused by two factors. On the one hand, the region selection strategy based on sliding window is not targeted, with high time complexity and window redundancy. On the other hand, the characteristics of manual design do not have good robustness to the variation of diversity. However, with the advent of target detection algorithm based on deep learning, target detection has made a great breakthrough. The breakthrough of R-CNN target detection algorithm lies in the use of Region Proposal + CNN instead of the traditional sliding window + manual design. In this way, target detection is decomposed into extraction of region proposal first, and then feature extraction + classification operation is carried out on the images within the region proposal.

The R-CNN algorithm uses the selective search to extract about 2000 region proposals which may contain objects from bottom to top in the image. The input is a color image, and the output is set of object location hypotheses. The selective search algorithm gives priority to merging four regions, namely those with similar color, similar texture, small combined total area and large proportion of combined total area in BBOX. These four rules only relate to the color histogram, texture histogram, area and location of the area. The combined regional features can be directly calculated from the sub-regional features. The selective search algorithm is shown in equation 3, 4, 5, 6, 7.

$$s_{color}(r_i, r_j) = \sum_{k=1}^{n} \min(c_i^k, c_j^k) \tag{3}$$

$$s_{texture}(r_i, r_j) = \sum_{k=1}^{n} \min(t_i^k, t_j^k) \tag{4}$$

$$s_{size}(r_i, r_j) = 1 - \frac{size(r_i) + size(r_j)}{size\ (im)} \tag{5}$$

$$\mathrm{fill}(r_i, r_j) = 1 - \frac{size(BB_{ij}) - size(r_i) - size(r_j)}{size\ (im)} \tag{6}$$

$$s(r_i, r_j) = a_1 s_{color}(r_i, r_j) + a_2 s_{texture}(r_i, r_j) + a_3 s_{size}(r_i, r_j) + a_4 \mathrm{fill}(r_i, r_j) \tag{7}$$

Although the selective search algorithm is no longer as exhaustive as the traditional target detection algorithm, the number of region proposals extracted by selective search is up to about 2000. All the suspected targets contained in these 2000 region proposals need CNN feature extraction and SVM classification. Therefore, the whole process requires a large amount of computation, which is also the reason for the slow detection speed of R-CNN algorithm.

However, it has been mentioned in the previous discussion that there are clear distinction between radar images and optical images. The low number of color channels in radar image is the most important feature. In addition, the regional division is also obvious in the radar image. Based on these features of radar images, an algorithm suitable for searching radar color block can be considered in region proposal determination of radar image. In this research, breadth-first search [15] (BFS) algorithm is proposed to replace the selective search algorithm. This is the improvement of R-CNN.

Breadth-first search is an image search algorithm. The algorithm starts at the root node at run time, traverses the nodes of the tree along the width of the tree, and terminates if a target is found. Breadth-first search is blind to the target search process, which can be understood as a blind search method. The purpose of the breadth-first search is to systematically expand and examine all the nodes in the diagram for results. In other words, the BFS algorithm thoroughly searches the entire graph until it finds a result, regardless of the possible location of the result.

The search process of breadth first algorithm is similar to the hierarchical traversal of tree. In operation, starting from a vertex in the graph, we first traverse each vertex, then all its adjacency points, and then from these adjacency points, we also visit their adjacency points successively. According to this process, the algorithm will not end until the adjacent points of all the accessed vertices in the graph are accessed.

The search process of breadth-first search algorithm is shown in Figure 6. In the figure, assume 1 as the starting point, and traverse all its adjacent points 2 and 3. Starting at 2, and traverse all its adjacent points 4 and 5. Starting at 3, and traverse all its adjacent points 6 and 7. Starting at 4, and traverse its adjacent point 8. Start at 5, and since all the starting points of 5 are already accessed, skip it. Points 6 and 7 are treated the same way as points 5.

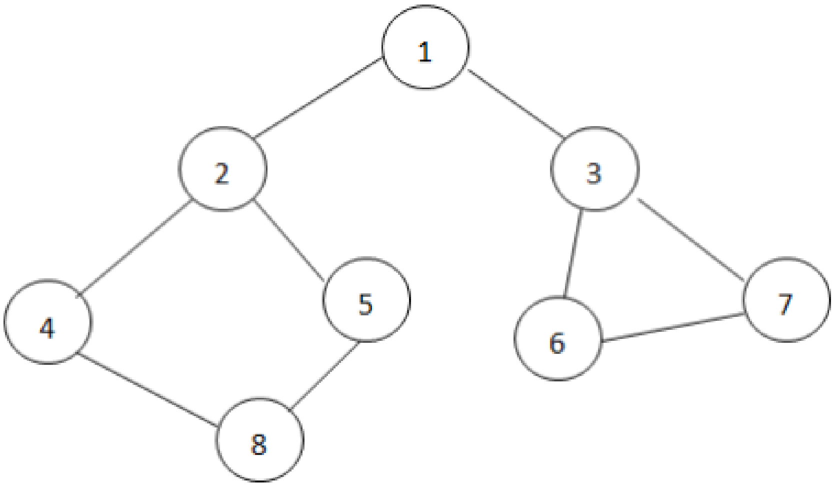


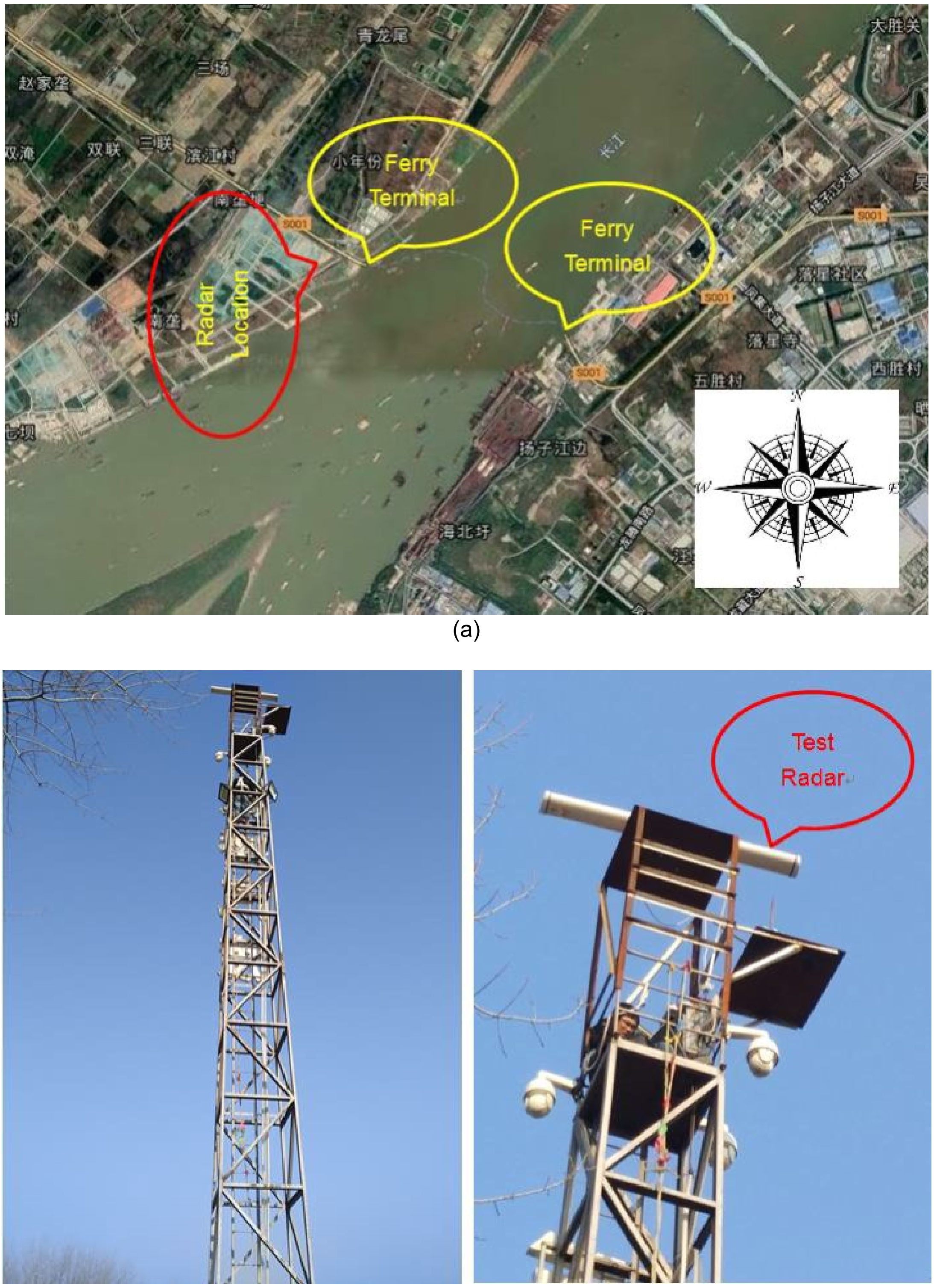
Figure 6: The search process of BFS

4 CASE STUDY

To validate the proposed approach, shore-based radar output images were selected as data. The radar images were taken from Banqiao ferry terminal, Nanjing City, Jiangsu province, China from 15:00 to 16:00 on 19th December 2017.

Experimental platform and process

The test water and experimental platform are shown in Figure 7. The experimental field is in Nanjing section of the Yangtze River Banqiao ferry terminal. The testing radar is Simrad HALO-6 pulse compression radar, and the radar detection range is 2 nm. The erection height of the radar is approximately 30m. The experimental duration is 60 minutes.



(a)

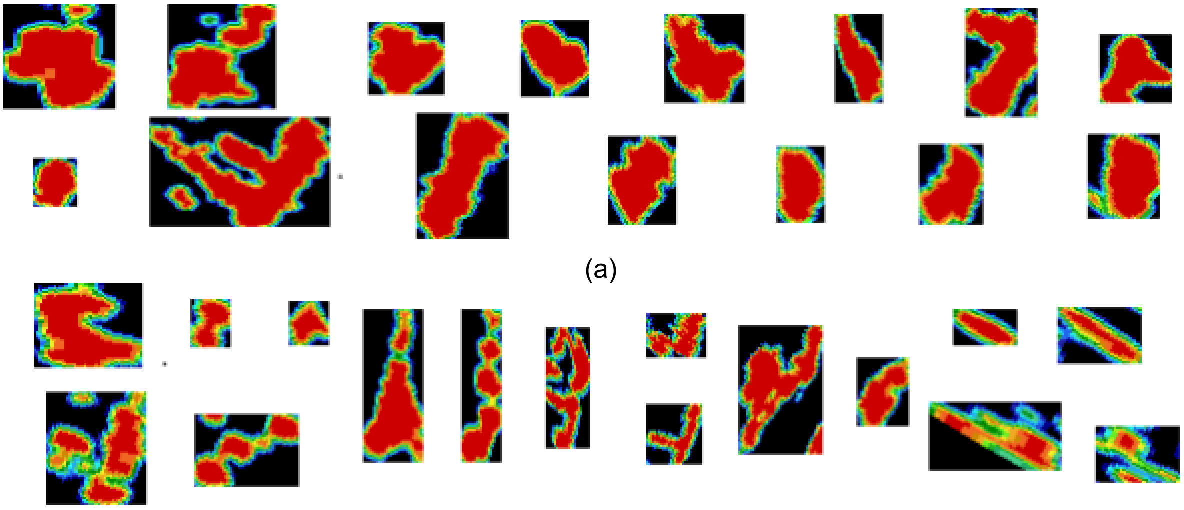
(b)

Figure 7: Radar image acquisition platform at the Banqiao ferry terminal, Nanjing, Jiangsu, China. Figure (a) is a satellite image of the test water. Figure (b) shows the experimental platform and test radar.

Step1: Feature extraction and SVM training

In this study, there are only two kinds of target detection in radar image: vessel and unknown object. Therefore, SVM dichotomy classifier is the best choice. The feature extraction for training dichotomy classifier model is completed by convolutional neural network. The target light spot detected in the radar image is obviously different from the background. Each target is independent of each other, and the target contour is obvious. Therefore, the complexity of CNN network used for feature extraction can be appropriately reduced. After verification, the network composed of two convolution layers, two pooling layers and relu is worked.

In addition, in general optical image target detection, there are a large number of samples for network training and testing, and it is extremely easy to obtain these samples. However, it is worth noting that what is presented in the radar image is a special representation of the perceived actual environment. This representation is fundamentally different from ordinary images. It is difficult for ordinary people to understand the meaning of radar images in a short time except for professional radar operators. There are few samples for training and testing. Therefore, a sample set needs to be created with the help of artificial experience. In the process of creating the sample set, in order to ensure a better feature extraction effect, a large number of negative sample sets should be established in addition to the positive sample set. A portion of the positive and negative sample set created by artificial experience is shown in Figure 8. Figure 8 (a) represents the ship spot sample, and Figure 8 (b) represents the unknown object spot sample.



(a)

(b)

Figure 8: Positive and negative sample set

Step 2: Radar image preprocessing

In the shore-based surveillance radar image, it is necessary to pay attention to the situation inside the channel, but not to put any effort on the situation outside the channel. The focus areas and parts that should be ignored in radar images are shown in Figure 9. In the image, the channel is in the white box, and the two sides of Nanjing section of the Yangtze River are outside the white box.

Moreover, the area outside the channel in the radar image also contains a large number of false signals and noise. When the algorithm processes the radar image, these useless signals will occupy too much resource and prolong the calculation time. It is necessary to filter out these unwanted signals. The radar image after using a mask to filter out unwanted signals is shown in figure 10.

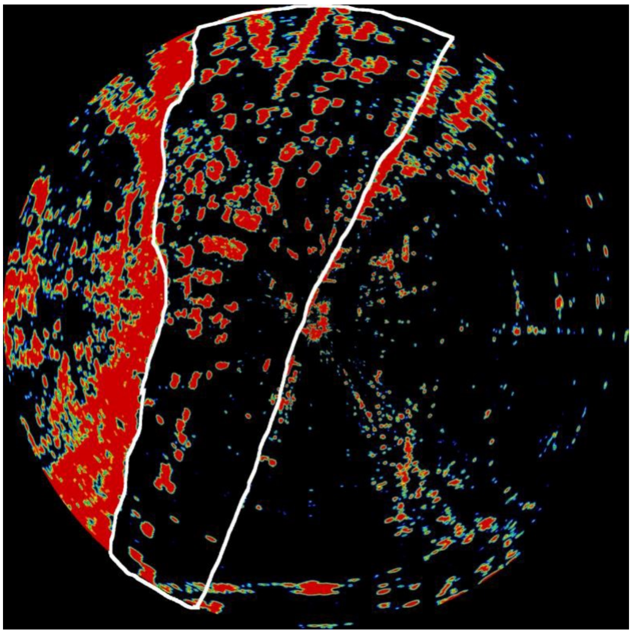


Figure 9: Radar images from the field

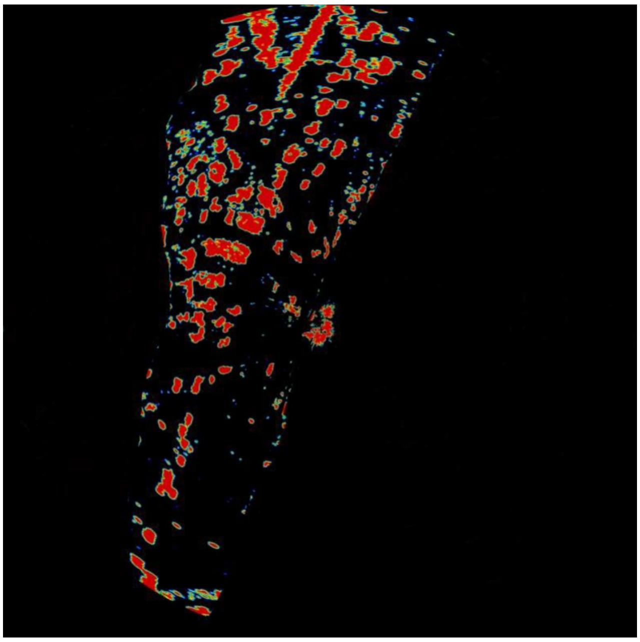


Figure 10: Preprocessed radar image

Step 3: Region proposal

The selective search algorithm extracts about 2,000 region proposals from the bottom to top in the image that may contain objects. Even with pre-processed radar images, the amount of computation is staggering. It still consumes a lot of computing time. In order to reduce the computation time, breadth-first search algorithm is proposed to replace selective search algorithm. The breadth-first search algorithm is more targeted in processing radar images. The reason is that the number of color channels in radar image is much smaller than that in optical image, and the spot area division of suspected target is very obvious, so breadth-first search algorithm can achieve very good results in terms of controllable time complexity in searching connected domain.

The specific process of the algorithm is as follows. In the pre-processed radar image, pixel points are selected within the channel area at a certain interval, and if the color conforms to the set threshold, the area will start to expand. The algorithm uses a queue to hold the outermost position, the outermost branch. Each branch has eight branches under it, and if one of them cannot continue to expand the region, the algorithm does not terminate, because there may be paths of other branches that can be expanded. The search is expanding in this way. The search ends when the queue is empty which means no branch in the outer location can continue to expand. A connected region is identified in the radar image.

Result analysis

In order to verify the detection effect of this study, video is also collected in the field while radar images are collected. Video was filmed at the same location as the radar on the north bank of the Yangtze River. Video screenshot is shown in Figure 11. After the radar images collected in the field are processed through the above steps, the detection results obtained are shown in Figure 12. In order to reflect the overall effect, the detection results in the channel are reintegrated with the outer filtering part of the channel. The yellow box represents the detection results of the ship, the blue box represents the other targets, and the error detection results are circled in red. Error detection results are determined by manual comparison of video images.

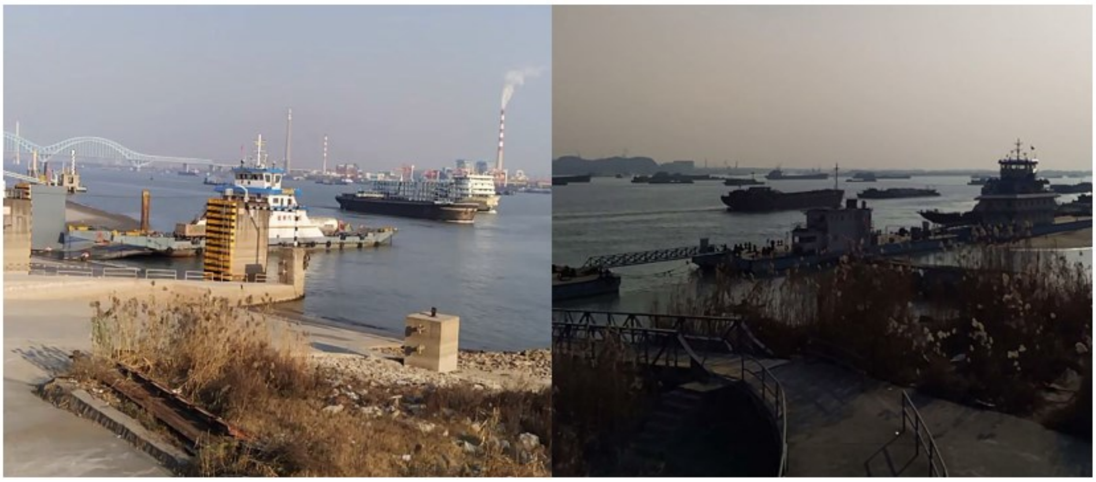


Figure 11: Video screenshot

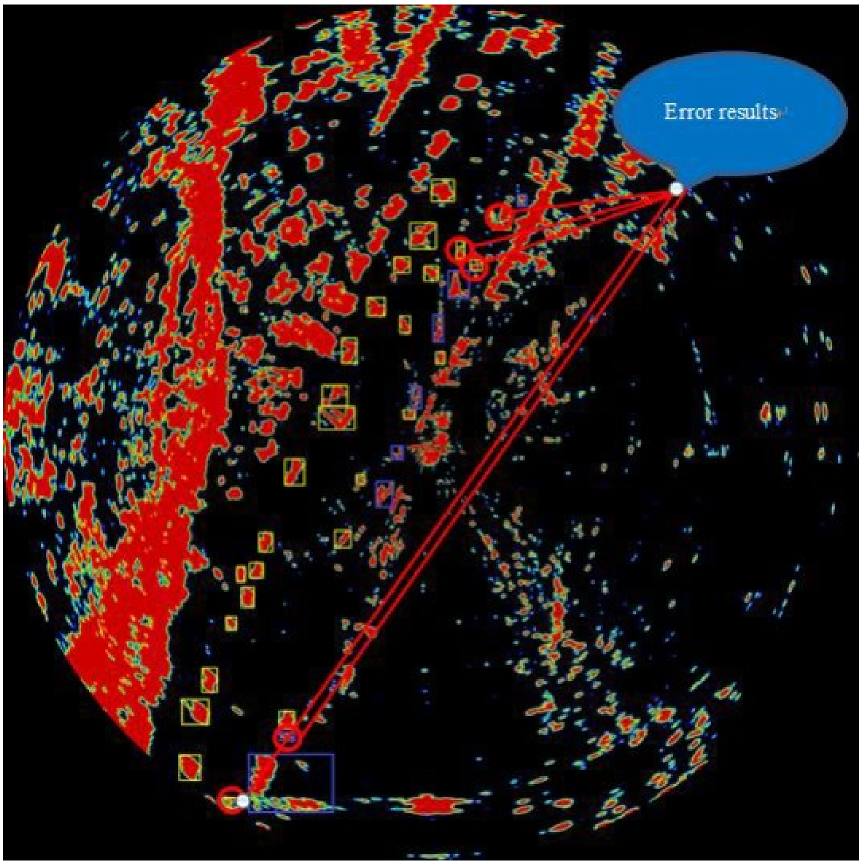


Figure 12: Detection result

The test statistical results are shown in table 1. According to the statistical results, a total of 35 targets were detected, among which 27 were vessels and 8 were other targets. By comparison with video images, there are 52 ships in video images that coincide with radar images, and the detection accuracy is 67.3%. There were four other targets wrongly judged as vessels, the detection accuracy is 85.2%. One vessel was wrongly judged as another target, the detection accuracy is 87.5%. In addition, according to the calculation results, the processing time of single frame radar image is 20 seconds, while the processing time of single frame image by R-CNN algorithm is about 47 seconds.

Table 1: Statistical table of detection result

	Vessel number	Vessel	Other object
Real quantity	52	23	7
Detection result	35	27	8
Ratio	67.3%	85.2%	87.5%

According to the obtained statistical data, the following conclusions can be drawn:

1. Target detection algorithm based on deep learning can detect targets in radar images.
2. The improved R-CNN algorithm proposed in this research can effectively detect vessels in the channel and has a good detection effect.
3. Through algorithm improvement, image processing time can be shortened on the original basis.

However, although the algorithm has been improved, there are still errors in the detection results. The misjudgment rate of real vessels was 14.8%, and that of other targets was 12.5%. The reasons for these errors are as follows:

1. The number of positive and negative sample sets for training and testing is not enough, so that the feature extraction network does not reach the optimal state.
2. Errors are caused by the working characteristics of the radar itself. Even the light spot contour of the same target will constantly change, which will have a certain impact on feature extraction.

5 CONCLUSION

As an important navigation aid in the shipping industry, marine radar has the ability to effectively perceive the environment in the navigation area and the situation in the channel. In the inland river crossing area, the output image of shore-based surveillance radar is capable of reflecting the position information of ferries and other vessels in real time, which plays an extremely important role in ensuring the navigation safety. At present, with the development of artificial intelligence, object detection based on deep learning can greatly improve efficiency and reduce labor cost. It is feasible to detect targets in radar image by deep learning algorithm. However, there is a big difference between radar image and optical image. Therefore, it is necessary to improve the algorithm when detecting targets in radar image.

In this research, R-CNN algorithm has been improved combined with the characteristics of radar image. Radar images are preprocessed to preserve only the channel area, which is capable reducing the amount of calculated data. The selective search algorithm of obtaining region proposals is replaced by the breadth-first search algorithm, which is capable further reducing the processing time. Through testing the radar images that collected in crossing area, the proposed approach was verified to be practical and the results were ideal.

However, due to the large amount of computation, the R-CNN algorithm is at disadvantage in computational speed, so that it cannot conduct real-time data processing. This obviously cannot satisfy the target detection of the real-time output radar image. In the next research, the research emphasis will be on algorithms with good real-time performance and detection accuracy, such as the Faster R-CNN

ACKNOWLEDGEMENTS

Thanks for the support from National Engineering Research Centre for Water Transport Safety and Nanjing Banqiao Ferry Company.

REFERENCES

[1] Girshick, R., Donahue, J., Darrelland, T., & Malik, J. (2014). Rich feature hierarchies for object detection and semantic segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE.

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition.

[3] Girshick, R. (2015). Fast R-CNN. IEEE International Conference on Computer Vision.

[4] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks.

[5] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You only look once: unified, real-time object detection.

[6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., & Fu, C. Y., et al. (2015). Ssd: single shot multibox detector.

[7] Viola, P. (2001). Rapid object detection using a boosted cascade of simple features. null. IEEE.

[8] Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE.

[9] Girshick, R. B., Felzenszwalb, P. F., & Mcallester, D. A. (2011). Object Detection with Grammar Models. International Conference on Neural Information Processing Systems. Curran Associates Inc.

[10] Masita, K. L., Hasan, A. N., & Paul, S. (2018, November). Pedestrian Detection Using R-CNN Object Detector. In 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI) (pp. 1-6). IEEE.

[11] Rujikietgumjorn, S., & Watcharapinchai, N. (2017, October). Vehicle detection with sub-class training using R-CNN for the UA-DETRAC benchmark. In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-5). IEEE.

[12] Chen, Y., Chen, J., Xiao, B., Wu, Z., Chi, Y., Xie, X., & Hua, X. (2019, April). Volume R-CNN: Unified Framework for CT Object Detection and Instance Segmentation. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) (pp. 872-876). IEEE.

[13] Ma, F. (2016). A novel marine radar targets extraction approach based on sequential images and bayesian network. Ocean Engineering, 120, 64-77.

[14] Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. NIPS (Vol.25). Curran Associates Inc.

[15] Najork, M., & Wiener, J. L. (2001). Breadth-first crawling yields high-quality pages.

Developing a Framework for Trustworthy Autonomous Maritime Systems

Dana Dghaym[1,*], Stephen Turnock[2], Michael Butler[1], Jon Downes[2], Thai Son Hoang[1] and Ben Pritchard[3]
[1] {d.dghaym t.s.hoang, mjb}@ecs.soton.ac.uk (Electronics and Computer Science, University of Southampton, UK)
[2] {s.r.turnock, j.j.downes}@soton.ac.uk (Maritime Robotics Lab, University of Southampton, UK)
[3] ben.pritchard@uk.thalesgroup.com (Thales, UK)

ABSTRACT

A key risk with autonomous systems (AS) is the trustworthiness of the decision-making and control mechanisms that replace human control. To be trustworthy, systems need to remain safe while being resilient to unpredictable changes, functional/operational failures and cybersecurity threats. Rigorous validation (does the solution satisfy the stakeholders' requirements and system's needs?) and verification (is the system free from errors?) are essential to ensure trustworthiness of AS. Current engineering practice relies heavily on Verification and Validation (V&V) test-and-fix of system characteristics which is very time-consuming and expensive, limiting the possibilities for exploration of alternatives in system design.

We present an approach to identifying and analysing mission requirements for squads of autonomous missions. Clear definition of requirements is an important pre-requisite for mission planning and for V&V of mission management. We use a structured approach to requirements identification and use formal modelling to help remove ambiguities in the requirements and to specify formal properties that should be satisfied by the missions. Our approach is being evaluated through consideration of a combined mission of the commercial C-Cat3 Unmanned Surface Vehicle (USV) (ASV Global, 2019) with deployment /recovery of small Unmanned Underwater Vehicles (UUV) within a shipping channel whereby the USV has to safely maintain station for a long period and then proceed to recover the UUV, while maintaining a communication link to an Unmanned Aerial Vehicle (UAV).

Keywords: Formal Methods; Event-B; Requirements; Maritime Autonomous Systems.

1. INTRODUCTION

Autonomous systems offer the potential of reducing the cost and ensuring the safety of humans. However, managing a squad of heterogeneous autonomous systems can be costly requiring a large number of people to complete a mission. This paper will focus on the early phases of designing an integrated mission management system for heterogenous autonomous assets, which we call the Integrated Mission Management System (IMMS). The aim of this system is to reduce the cost of missions requiring multiple platforms.

Mission management involves the following activities: planning of a mission after identifying the mission goals, mission execution and reviewing the mission. While we have trustworthy autonomous vehicles working as separate entities, our aim is to build a trustworthy management system to ensure the trustworthiness of the overall system.

---

[*] d.dghaym@ecs.soton.ac.uk

Studies have shown that the cost of fixing errors during testing is 10 times more than during the construction phase and can increase to more than 25 times post release (Leffingwell, 1997), and many problems discovered in software systems are related to shortcomings in requirements elicitation and specification processes (MacDonell et al. , 2014). In this paper, we will show how we can apply formal modelling to develop a requirements analysis framework for identification of anticipated range of operational environments for autonomous missions, including human operator interactions, together with precise specification of safety and security envelopes for enactment of autonomous missions.

This paper is organised as follows: Section 2 presents some background information about Event-B formal method. Section 3 gives an overview of the approach followed in identifying and analysing the mission requirements. Section 4 describes how we apply formal modelling to identify the system requirements. Finally we present our conclusions in Section 5.

2. BACKGROUND

In this section, we present Event-B, a formal method for system analysis and modelling. We have chosen Event-B because Event-B supports modelling at a system level rather than only at a software level. Event-B also has a good extensible tool support and a user can apply both theorem proving and model checking, supported by ProB (Leuschel & Butler, 2008), to the same model. A survey of formal verification tools have found that Event-B supported by the toolset Rodin comes closest to supporting the goals of a *correct-by-construction* designs (Armstrong et al. , 2014). In this paper, we use Event-B to address the ambiguity and inaccuracy of requirements specifications.

2.1 Event-B

Event-B (Abrial, 2010) is a formal method for system development. One of the main features of Event-B is the use of ***refinement*** to introduce system details gradually into the formal model. An Event-B model consists of two parts: ***contexts*** and ***machines***. Contexts are the static parts of the model. A Context contains ***carrier sets***, ***constants***, and ***axioms*** that constrain the carrier sets and constants. Machines are the dynamic parts of the model. A machine contains ***variables*** v, ***invariants*** $I(v)$ that constrain the variables, and ***events***. An event comprises a guard denoting its enabling-condition and an action describing how the variables are modified when the event is executed. In general, an event e has the following form, where t are the event parameters, $G(t, v)$ is the guard of the event, and $v := E(t, v)$ is the action of the event.

e == **any** t **where** $G(t,v)$ **then** $v := E(t,v)$ **end**

A machine in Event-B corresponds to a transition system where ***variables*** represent the states and ***events*** specify the transitions. Contexts can be ***extended*** by adding new carrier sets, constants, axioms, and theorems. Machine M can be ***refined*** by machine N (we call M the abstract machine and N the concrete machine). The state of M and N are related by a gluing invariant $J(v, w)$ where v, w are variables of M and N, respectively. Intuitively, any "behaviour" exhibited by N can be simulated by M, with respect to the gluing invariant J. Refinement in Event-B is reasoned event-wise. Consider an abstract event e and the corresponding concrete event f. Somewhat simplifying, we say that e is refined by f if f's guard is stronger than that of e and f's action can be simulated by e's action, taking into account the gluing invariant J. More information about Event-B can be found in (Hoang, 2013). Event-B is supported by Rodin (Abrial et al., 2010), an extensible toolkit which includes facilities for modelling, verifying the consistency of models using theorem proving and model checking techniques, and validating models with simulation-based approaches.

3. AN APPROACH FOR REQUIREMENTS ELICITATION

Defining the requirements of the IMMS is an iterative process, where in the initial version we focus on ***what we know***. We start by gathering information about the different available autonomous platforms. In this case, we have three different physical assets from each of the three domains: surface, underwater and aerial. After defining the system goals, assumptions and constraints, which include communication and planning constraints in addition to identifying the failure and adverse conditions, we structure the requirements as follows:

1. Operator Safety Requirements
2. Platform Functional Requirements
    a. Unmanned Surface Vehicle (USV)
    b. Unmanned Underwater Vehicle (UUV)
    c. Unmanned Aerial Vehicle (UAV)
3. Possible Exceptions and Recovery Actions
4. Security Requirements

In the initial version, our focus is on the available assets and their interfaces to the IMMS. After analysing the existing requirements, we identify what is missing, in this case it is clearly the IMMS requirements or in other words ***what we want***. From what we know and what we want, we identify the functional and non-functional requirements of the IMMS, the IMMS interface requirements and information communication. The functional requirements include mission planning, mission execution and mission monitoring and review. Later, we can identify a common functionality among the different assets and generalise the platform-specific requirements.

Capturing the requirements in a well-defined document is not enough. The document can be still prone to different interpretations from the different team members coming from different backgrounds. Therefore, it is important to have a precise specification to eliminate any ambiguities and remove any defects. For this we use Event-B, introduced in Section 2.1, to capture the system requirements precisely. In Section 4 we present an early attempt at modelling the high level requirements of the system using Event-B.

Figure 1 presents our proposed approach for eliciting requirements for autonomous missions. This approach is based on our experience in using formal modelling for system verification, it is a generic approach which is applicable for the integration of multi-platforms. Our approach is iterative where we augment these requirements as a result of continuous analysis. This approach requires continuous interaction between two main stakeholders the domain experts, which in our case includes two parties: the different platform owners and the client (Thales), and the formal methods experts. The platform owners will identify what are the feasible requirements and the client identifies what is the purpose of the system. The formal methods experts will work on analysing the available resources to identify what is missing and what is ambiguous which will need clarification from the domain experts, who should also approve or reject any identified requirements. In the next sections we apply this approach to defining the requirements of the IMMS.

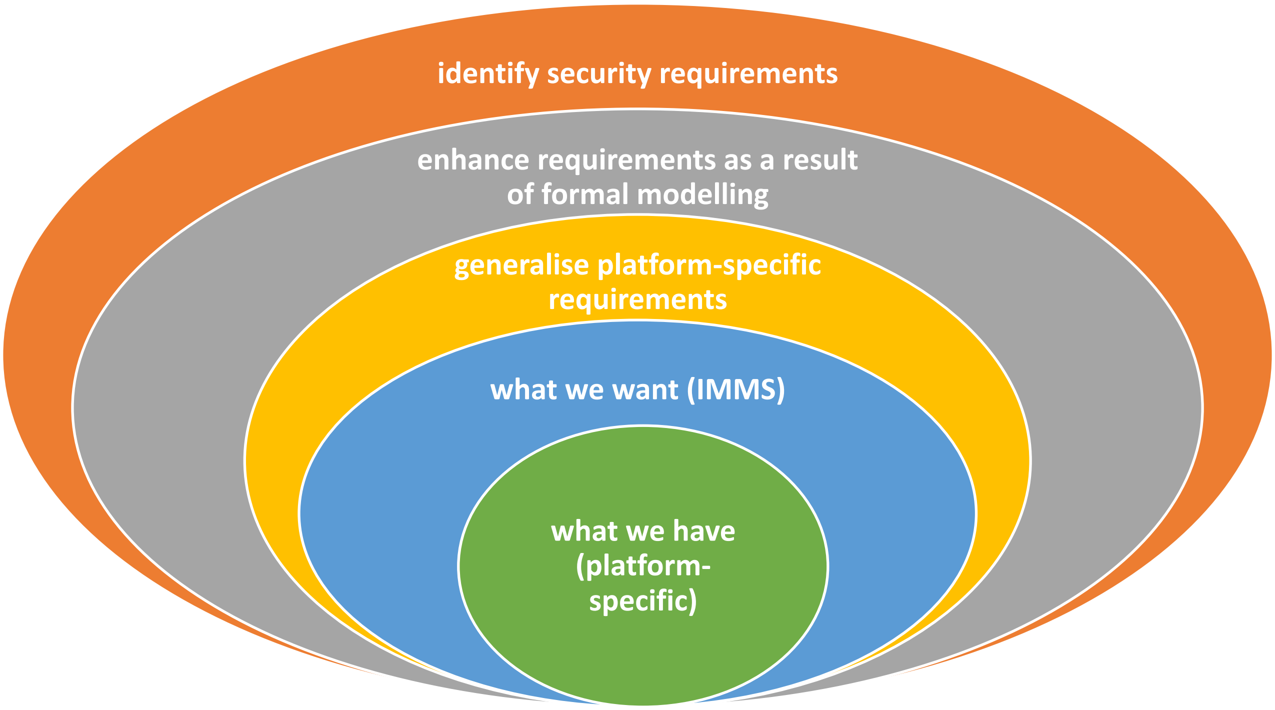

identify security requirements
enhance requirements as a result of formal modelling
generalise platform-specific requirements
what we want (IMMS)
what we have (platform-specific)

Figure 1: An approach for eliciting requirements

3.1 Analysis of an IMMS Safety Requirement

In this section, we will focus on one of the safety requirements of the IMMS, which we use as a running example to illustrate our approach rather than presenting the full set of requirements. In the first stage we looked at identifying the operator Safety Requirements (SF) of the available assets, one of these requirements is:

SF1. Collision Avoidance (CA): The UxVs (unspecified unmanned vehicles) do not have collision avoidance mechanisms. Collision avoidance with other vehicles and other possible obstacles is maintained by thorough planning, which can be updated during the mission should conditions change. Additionally, maintaining *visual line of sight*, receiving video feeds from platforms and defining mitigation scenarios assist with collision avoidance.

CA Requirement Analysis: The available vehicles do not have collision avoidance mechanism. Therefore, in order to avoid collisions:

A. Initial planning should take into considerations the different assets positions and any known obstacles in the environment.
B. During the mission when situational awareness is available and the assets are communicating, the plans can be updated and sent to the assets to avoid collisions.
C. A *timeout* should be predetermined for the assets with a predetermined plan to follow in case of communication loss.

Identifying IMMS functional requirements: By analysing the **SF1,** we can identify some of the IMMS Functional Requirements (FR) which **should** include:

FR1. The IMMS must have the ability to specify/assign the required vehicles to perform a mission.
FR2. The IMMS must assign tasks to the specified vehicles.
FR3. The IMMS must provide the vehicles with initial plans prior to starting a mission.
FR4. The IMMS must have the ability to modify plans of assigned vehicles during the mission executions.

Both **FR1** and **FR2** can be inferred from **A**, since planning should know the positions of the mission assets, then it should have the ability to assign these assets to a mission and give them tasks to perform a mission. From both **A** and **C**, **FR3** is deduced which will result in providing the vehicles with plans in the case of normal and failure behaviours. **FR4** is clearly concluded from **B** where plans should be updated should problems arise.

In this section, we have shown how we can identify some of the system requirements by analysing the requirements of *what we know.*

4. MODELLING IN EVENT-B

A key strength of Event-B is refinement, which allows us to abstract away from details and focus on different problems at different levels of refinements. The goal of this early modelling is an attempt to understand the system under development, remove ambiguities and identify important missing properties of the system to enhance the requirements.

Our Event-B model starts with an abstract level defining a mission as a set of tasks. Then, we introduce two refined levels as: vehicles and mission planning.

4.1 Abstract Level: Mission

At this level we define a mission as a set of tasks and introduce the events: *define_mission*, *start_mission* and *complete_mission*. These events will execute in the following order: <*define_mission*; *start_mission*; *complete_mission* >.

This level has three simple events, however right at the start of modelling we have to take a modelling decision: *Can the IMMS manage multiple simultaneous missions*?

For this project we will manage one mission at a time and leave this question as a future research question. Other questions that we have identified at this level are as follows:

- *What are the conditions for starting a mission, or we can ask, do we need all the vehicles to be present to start a mission?*
- *What are the conditions for completing a mission?*

We can define a new requirement for the IMMS, related to starting a mission:

FR5. The IMMS should define a minimum criterion for starting a mission.

In the following section, we will show how we can model this FR5 requirement by introducing a new refinement level.

4.2 First Refinement: Vehicles

In the first refinement level, we introduce vehicles and their capabilities and the possibility of assigning vehicles to mission tasks. Figure 2 represents a class diagram of the static part of the model, Vehicles context, while Figure 3 represents a class diagram of the dynamic part of the vehicles model. This is a UML-like representation of the model, with a formal translation to Event-B called UML-B class diagram (Snook & Butler, 2008) (Snook & Butler, 2003). This class diagram shows the different relationships between the different classes of the model and some of the class methods which are translated to events in the Event-B machine. In the context, the classes are translated to either sets or constants, in our case sets, while in the machine they are translated to variables but the machine can still reference the context classes as shown in Figure 3. The associations are translated to constants in the context and to variables in an Event-B machine.

The main functionality at this level is to ensure that a mission can only start after assigning vehicles with the minimum required capabilities defined to start the mission tasks. In Event-B this is ensured by defining the following invariant which must be maintained by all the events:

```
@inv1: missionStart = TRUE ⇒ requiresMin[missionTasks] ⊆ capabilities[assign∼[missionTasks]]
```

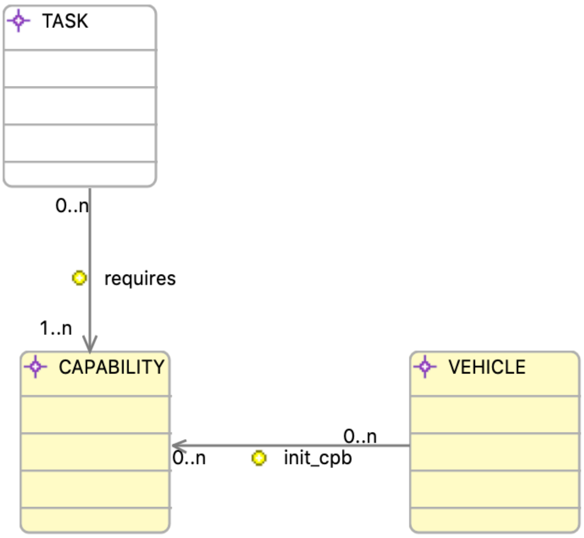


Figure 2: Vehicles Context

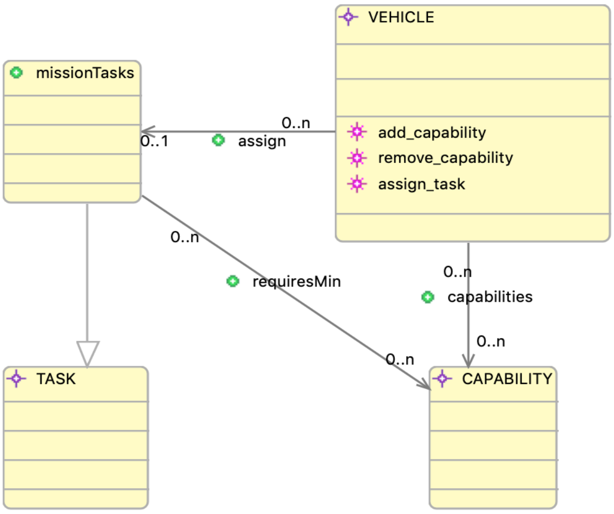


Figure 3: Vehicles Machine

In Event-B, proof obligations will be generated to ensure that all events will maintain the defined invariants. Invariant ***inv1*** will address some of the requirements, in this case requirement **FR5**. In addition to that Event-B can help to prove the consistency of the invariants, for example if we have additional invariants that conflict with each other, it will be impossible to prove the model and hence it will flag a problem to the modeller and requirements can be changed accordingly.

However, when starting this early we came up with additional questions that were not defined clearly in the requirements document, for example: *Do we allow vehicles to be deallocated from their tasks before mission completion?* If yes, then this invariant cannot be maintained by all the events, however we can address the minimum requirement of starting a mission as a guard in the event *start_mission*, which is a precondition to execute the event, but is not necessarily maintained by all the other events.

In this section, we have shown an example of how we can use Event-B to improve the requirements and identify some defects. We will also show how to trace the requirements in the model in Section 4.6. In this case, some invariants and guards are added to address some requirements for starting a mission that is why it is important to label the requirements to facilitate their traceability in the model.

4.3 Second Refinement: Mission Planning

At this level, we introduce mission plans abstractly as a series of locations, and in our model we ensure that a mission is not considered successfully complete until all vehicle plans are covered. We also introduce an event to set an initial plan before starting a mission and another event that enable modifying plans during execution, addressing requirements **FR.3** and **FR.4**.

This level also poses new questions about the conditions for modifying plans, is it always a response to some changes to the environment, do we immediately modify the plan or does the vehicle has to go through a safe state?

To answer these questions, we suggest defining generic states that apply to all the vehicles and define 'what are the activities that can occur during these states?'. These activities and the state transitions will be modelled as events in Event-B. A possible generic state-machine for the vehicles is shown in Figure 4.

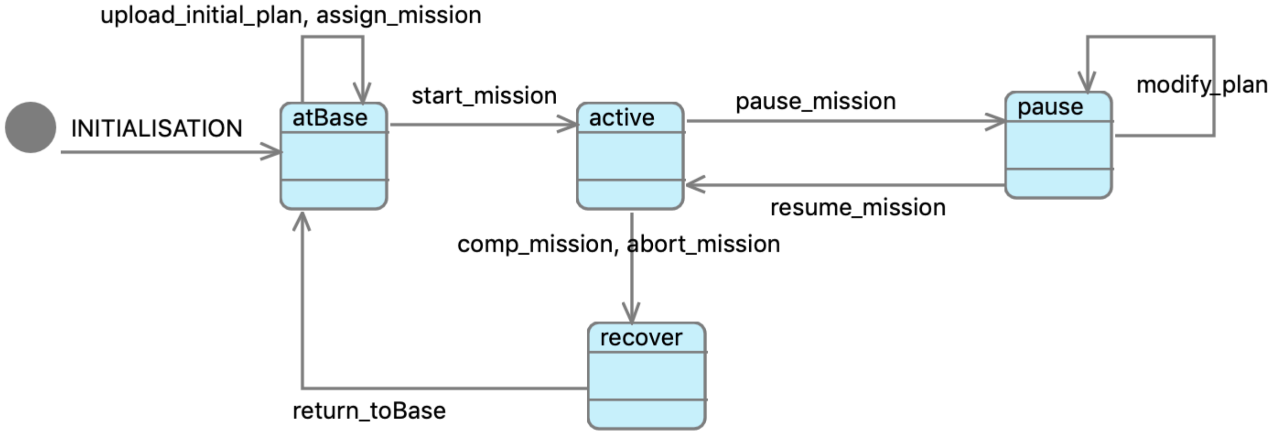


Figure 4: Generic Vehicle State machine

We can then refine the generic states by defining the tasks specific to each vehicle. For example, in an active state a USV will do the following events in order: move_to_survey_area, deploy_UUV, then recover_UUV. Similarly we can define another vehicle-specific events for a UUV such as: dive, survey and then surface.

4.5 Future Refinements & Security Requirements

In the previous sections, we have defined a high level abstraction of the IMMS, and we have shown how we used Event-B to identify new requirements, discover some ambiguities that require decisions from domain experts. After enhancing the requirements based on our formal modelling, we will introduce timing constraints, and model any remaining requirements by introducing new refinements. We will also extend the contexts to provide some instantiations to better fit with the mission types and provide a basis for mission validations.

In this paper, we have not described the security requirements, but in previous work (Omitola et al, 2018) and (Omitola, Rezazadeh, & Butler, 2019) we have used System Theoretic Process Analysis (STPA) (Leveson & Thomas, 2018) and STPA-Sec (Young & Leveson, 2013) to analyse the security requirements of maritime cyber-physical systems. In (Snook, Hoang, & Butler, 2017), we propose a general approach based on abstraction and refinement to analyse and construct security protocols using Event-B together with UML-B class diagrams and state-machines for diagrammatic visualisations. Regarding the IMMS security requirements, we intend to follow a similar approach using STPA to analyse, identify the unsecure scenarios and define the mitigation scenarios and constraints, then use Event-B and UML-B for verifying the system constraints.

4.6 Requirements Traceability in Event-B

In this section we show in detail how we can capture a requirement in Event-B. Table 1 uses requirement **FR.5** presented in Section 4.1 as an example. We show how existing events are extended with new guards and actions to capture the requirement. This requirement is enforced by an invariant. The importance of the invariant is to ensure that the defined constrained is maintained by all the events.

Table 1: Requirement Traceability

Req. ID	Model	Representation in the model	Event-B Syntax
FR. 5	1st Refinement: Vehicles (Machine)	*define_mission*: Action to set minimum required capability	requiresMin ≔ t × cpb
		start_mission: Guard to check that the minimum required capabilities is already assigned to the mission tasks.	requiresMin[missionTasks] ⊆ capabilities[assign∼[missionTasks]]
		Invariant to ensure that a mission can only start if the required minimum capabilities are assigned (inv1).	missionStart = TRUE ⇒ requiresMin[missionTasks] ⊆ capabilities[assign∼[missionTasks]]

For example, we have an event *remove_capability* that allows us to remove a capability from a vehicle, if this event is not constrained, Event-B will not be able to discharge the proof obligation related to maintaining invariant (***inv1***) in this event, hence the modeller will discover that something need to be changed, in this case we need to add a guard that prevents removing a capability from a vehicle with assigned task.

5. CONCLUSIONS

Autonomous systems are safety-critical systems, hence the need for assurance techniques is a necessity for trusting such systems and be certified for use. Having a trustworthy part of the system is not enough to trust the trustworthiness of the overall system and having a heterogeneous system that involve platforms from different domains adds another level of complexity to gain certification. Maritime operating environments are particularly challenging for V&V. The environmental conditions can cause vehicle failures or impede communications i.e., interconnecting autonomous systems, could potentially open up these systems to more security attacks. Unpredictability of the maritime environment can mean that plans need to be updated autonomously during mission execution. It may not be feasible to characterise, fully, the environmental conditions and required system responses of AS in advance of deployment. In future work, we aim to be able to characterise the ***safety and security envelopes*** within which system responses should reside. An emerging approach to ensuring safety in Artificial Intelligence (AI)-based systems is to augment them with ***policing functions*** (Hoang et al, 2018) that monitor AI decision-making for conformance to safety/security envelopes so that, when an unsafe/insecure decision is detected, some failsafe action is invoked, e.g., command a drone to loiter or vessel to surface. Ideally, safety/security envelopes can be characterised in precise ways, making sure policing functions are amenable to the characterisation required for assurance cases in advance of deployment.

In this paper, we have proposed an approach for eliciting requirements for autonomous missions and formalising these as Event-B models. This is part of the functional process for an Integrated Mission Management for heterogenous autonomous systems. Figure 1 summarises the proposed approach and shows how we augment the requirements through continuous analysis. The proposed approach is iterative where we continuously need to do reviews which can influence the formal modelling on one hand and the formal modelling can influence the system requirements by identifying new requirements, removing ambiguities and defects. At the early stages we are using formal modelling for requirements and design analysis and to prove the consistency of the system properties. Later we could use formal modelling to verify and validate the high level plans.

ACKNOWLEDGEMENTS

This work is funded by Thales IMMS project.

REFERENCES

Abrial, J.-R. (2010). *Modeling in Event-B: System and Software Engineering*. Cambridge University Press.

Abrial, J.-R., Butler, M., Hallerstede, S., Hoang, T. S., Mehta, F., & Voisin, L. (2010). Rodin: an open toolset for modelling and reasoning in Event-B. *International Journal on Software Tools for Technology Transfer*, *12*(6), 447–466. https://doi.org/10.1007/s10009-010-0145-y

Armstrong, R. C., Punnoose, R. J., Wong, M. H., & Mayo, J. R. (2014). *Survey of Existing Tools for Formal Verification*. https://doi.org/doi:10.2172/1166644

Hoang, T. S. (2013). *An Introduction to the Event-B Modelling*. https://doi.org/10.1007/978-3-642-33170-1

Hoang, T. S., Sato, N., Myojin, T., & Butler, M. (2018). Policing Functions for Machine Learning Systems. In *Workshop on Verification and Validation of Autonomous Systems: Satellite Workshop of Floc 2018* (pp. 1–10). Retrieved from https://eprints.soton.ac.uk/421233/%7D

Leffingwell, D. (1997). Calculating the Return Investment from more Effective Requirements Management. *American Programmer, 10*(4), 13–16.

Leuschel, M., & Butler, M. (2008). {ProB}: An Automated Analysis Toolset for the {B} Method. *Software Tools for Technology Transfer (STTT)*, *10*(2), 185–203.

Leveson, Nancy G. and Thomas, J. P. (2018). *STPA Handbook*. Cambridge, MA USA.

MacDonell, S. G., Min, K., & Connor, A. M. (2014). Autonomous requirements specification processing using natural language processing. *CoRR*, *abs/1407.6099*. Retrieved from http://arxiv.org/abs/1407.6099

Omitola, T., Downes, J., Wills, G., Zwolinski, M., & Butler, M. (2018). Securing Navigation of Unmanned Maritime Systems. In N. C. Schillai, Sophia M. and Townsend (Ed.), *Proceedings of the 11th International Robotic Sailing Conference: Southampton, United Kingdom, August 31st - September 1st, 2018.* (pp. 53–62). Southampton: CEUR-WS. Retrieved from http://ceur-ws.org/Vol-2331/paper5.pdf

Omitola, T., Rezazadeh, A., & Butler, M. (2019). Making (Implicit) Security Requirements Explicit for Cyber-Physical Systems : A Maritime Use Case Security Analysis. In G. A.-K. et Al (Ed.), *Proceedings of the 30th DEXA Conferences and Workshops*. Linz, Austria: Springer Nature Switzerland AG 2019. https://doi.org/https://doi.org/10.1007/978-3-030-27684-3_11

Snook, C., & Butler, M. (2003). UML-B : Formal modelling and design aided by UML, 1–32.

Snook, C., & Butler, M. (2008). UML-B and Event-B: an integration of languages and tools. In *The IASTED International Conference on Software Engineering - SE2008*. Retrieved from https://eprints.soton.ac.uk/264926/

Snook, C., Hoang, T. S., & Butler, M. (2017). Analysing Security Protocols Using Refinement in iUML-B. In C. Barrett, M. Davies, & T. Kahsai (Eds.), *NASA Formal Methods* (pp. 84–98). Cham: Springer International Publishing.

Young, W., & Leveson, N. (2013). Systems Thinking for Safety and Security. In *Proceedings of the*

29th Annual Computer Security Applications Conference (pp. 1–8). New York, NY, USA: ACM. https://doi.org/10.1145/2523649.2530277

International Seminar on Safety and Security of Autonomous Vessels
17 - 18 September 2019, Helsinki

Aalto University

Korean Technical Innovation: toward Autonomous Ship and Smart Shipbuilding to Ensure Safety

Yongwon Kwon[1,*], Yong-kuk Jeong[2], Jong-hun Woo[3], Daekyun Oh[4], Hyunwoo Kim[5], Il-Sik Shin[5], Seungho Jung[1], Sanghyuk Im[1], and Changho Jung[1]

[1] Research Institute of Medium & Small Shipbuilding (Department of Strategic Planning, South Korea)
[2] KTH Royal Institute of Technology (Department of Sustainable Production Development, Sweden)
[3] Seoul National University (Department of Naval Architecture and Ocean Engineering, South Korea)
[4] Mokpo National Maritime University (Department of Naval Architecture and Ocean Engineering, South Korea)
[5] Research Institute of Medium & Small Shipbuilding (Department of ICT & Smart Ship, South Korea)

ABSTRACT

Recently, the fourth industrial revolution has resulted in a paradigm shift to the development of information and communication technology and smart technology, which has extended to the shipbuilding industry worldwide. The Korean shipbuilding industry employs a high level of smart application technology, smart production technology, and smart ship platforms, which have been extended to all areas in the value chain, resulting in relatively high added value. The project “The Smart K-Yard” was conceived to achieve growth in the productivity of Korean small and medium-sized shipyards. This project is necessary to eliminate elements of waste and minimize the work force, thereby reducing the production lead-time and enhancing the quality of intelligent and smart shipbuilding production system. The project uses automation technology and simulation-based test bed for the shipbuilding processes. It has five work packages: 1) intelligent shipbuilding production design platform, 2) shipbuilding process automation technology, 3) shipyard operational efficiency improvement technology, 4) simulation-based virtual production platform, and 5) Smart K-yard supporting and training center using digital twin and cyber-physical system to improve and maintain the level of skills.

In this paper, the development of unmanned vessels control system as part of future ships in South Korea is introduced. The system is divided into the hull, propulsion system, steering system, control system, and power system. The main controller for the engine, waterjet, and power has been developed. Moreover, the reliability and usefulness of the systems were verified using test beds and water tank testing. We plan to continue the development of advanced technology for unmanned vessel, including determining the appropriate safety improvement method to be implemented in the event of damage caused by abnormal condition and operator’s negligence during operation and maintenance of equipment as an unmanned vessel is equipped with more communication related devices and electronic equipment than a manned vessel.

Technological cooperation for future ships focusing on industrial internet of things (IIoT) sensor technology, robotic system, and unmanned vessels is increasingly necessary among different countries. As such, we highlighted the need to increase information technology cooperation between research institutes, universities, and relevant companies in Finland in response to the fourth industrial revolution of shipbuilding industry.

Keywords: Autonomous Vessels; Smart Ship; Smart Shipbuilding; Digital Twin; Cyber Physical System

1. INTRODUCTION

In order to secure price competitiveness, the Korean shipbuilding industry is striving to upgrade the shipbuilding ecosystem by linking the paradigm shift in intelligent smart production

---

[*] Corresponding author: +82-51-974-5529, navykwon04@gmail.com

concepts in the shipbuilding industry and the overall value chain with the supplier and shipbuilders.

Moreover, small and medium sized shipyards have limitations in human resources and capital capacity compared with large shipyards. Consequently, their competitiveness is gradually weakened in terms of specialized technical skills to achieve future goals due to increase in the technological gap. To secure competitiveness in the shipbuilding process, small and medium sized shipyards need to improve their planning and management capabilities and develop core technologies for future shipbuilding process.

Korean shipbuilding companies are seeking to reduce construction costs by upgrading their shipbuilding technology through automated process management, production automation, and big data. Moreover, they are proposing a smart shipyard business model to enhance productivity and cope with low-cost orders, increased global competition, and oversupply.

The development of information service support system, such as the establishment of optimal wired and wireless communication infrastructure within the world's first commercialized 5G-based shipyard, is needed to maximize productivity and cost savings. This is also useful for the provision of knowledge-based optimal production management services using big data technology, as well as real-time sharing and integrated management of information among workers.

In addition, technological advancement in the global shipbuilding industry over the past few years have resulted in increased effort to improve the performance of ships and secure technological competitiveness while establishing the concepts of smart ships/intelligent ships using new technologies such as digital twin to enhance the safety of ships and reduce operating costs. Furthermore, significant progress has been made in the automobile industry as a result of the fourth industrial revolution, such as self-driving cars, smart cars, and connected cars.

In addition to the shipbuilding industry, various technologies such as the internet of things (IoT), big data analytics, cyber security, simulation, remote maintenance, real-time monitoring, and integrated control are also required. Although minimal cooperation exists between the shipbuilding industry and the shipping industry, cooperation with the shipping industry has been recently expanded in the form of smart ships considering operational efficiency and safety.

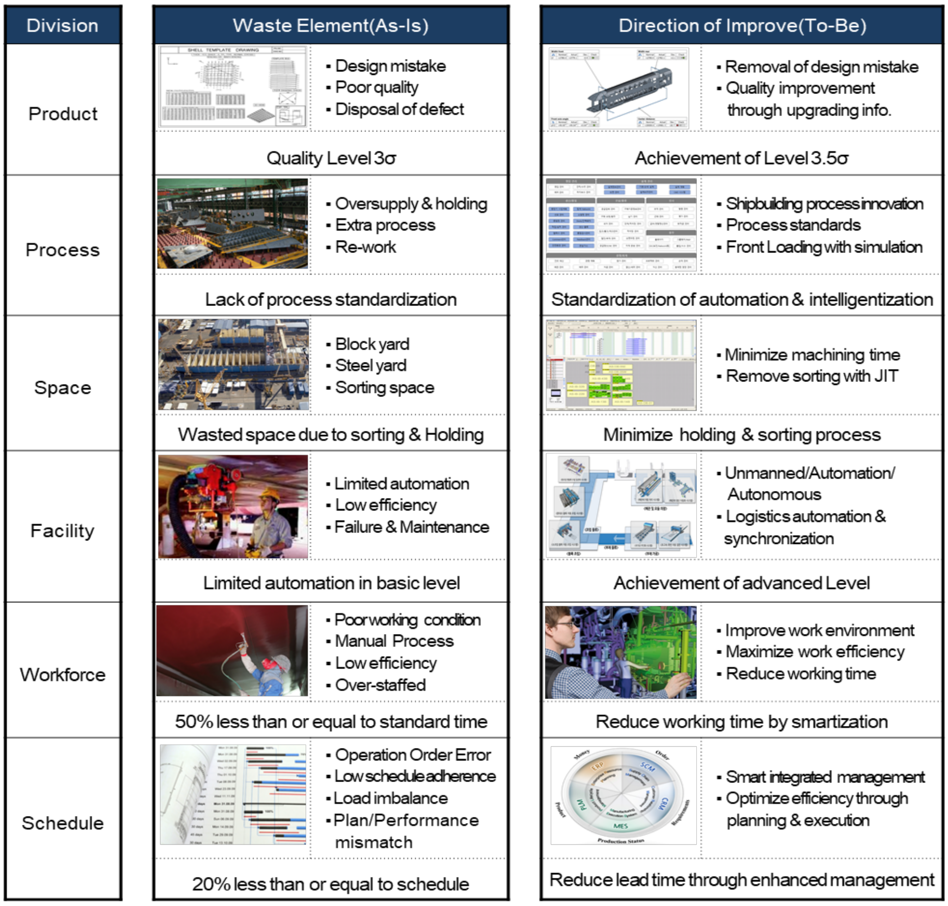
However, there are high restrictions, relatively high communication costs compared with land-based systems, and difficulty in securing connectivity due to slow communication speed in remote sea communication environment, unlike self-driving cars in land-based wireless communication environment.

In addition, maritime transport should comply with strict standards from the International Maritime Organization (IMO); these international standards and regulations are a relatively slow factor for derived solutions and services to be commercialized. To address this problem, we proposed plans to secure the competitiveness of Korean shipbuilding companies through a smart shipbuilding process and Korean small and medium-sized autonomous ships.

2. THE SMART K-YARD

2.1. Component of the Smart K-Yard

Smart shipyards are defined as ship production systems that can combine the latest information and communication technology (ICT) and automation technology to eliminate waste elements from products, processes, schedules, space, facilities, and human resources. A smart shipyard also optimizes energy operations, reduces production lead time, and ensures quality. It uses a smart shipyard operating system to optimize various complex materials, parts, and processes through simulation engineering. Figure 1 shows the components and strategic plan of the Smart K-Yard.



Division	Waste Element(As-Is)	Direction of Improve(To-Be)
Product	▪ Design mistake ▪ Poor quality ▪ Disposal of defect Quality Level 3σ	▪ Removal of design mistake ▪ Quality improvement through upgrading info. Achievement of Level 3.5σ
Process	▪ Oversupply & holding ▪ Extra process ▪ Re-work Lack of process standardization	▪ Shipbuilding process innovation ▪ Process standards ▪ Front Loading with simulation Standardization of automation & intelligentization
Space	▪ Block yard ▪ Steel yard ▪ Sorting space Wasted space due to sorting & Holding	▪ Minimize machining time ▪ Remove sorting with JIT Minimize holding & sorting process
Facility	▪ Limited automation ▪ Low efficiency ▪ Failure & Maintenance Limited automation in basic level	▪ Unmanned/Automation/ Autonomous ▪ Logistics automation & synchronization Achievement of advanced Level
Workforce	▪ Poor working condition ▪ Manual Process ▪ Low efficiency ▪ Over-staffed 50% less than or equal to standard time	▪ Improve work environment ▪ Maximize work efficiency ▪ Reduce working time Reduce working time by smartization
Schedule	▪ Operation Order Error ▪ Low schedule adherence ▪ Load imbalance ▪ Plan/Performance mismatch 20% less than or equal to schedule	▪ Smart integrated management ▪ Optimize efficiency through planning & execution Reduce lead time through enhanced management

Figure 1: Components and strategic plan of Smart K-Yard

2.2. Smart shipyard assessment model

To transform a shipyard into a smart shipyard, the concepts of lights out factory for automated production systems and information systems, and connected factory for connecting shipyard products and resources should be employed. Digital twin, machine learning, and cyber-physical system (CPS) are also important to maximize production capacity. In addition, shipyard-focused enterprise management systems such as ERP, MES, SCM, and APS should be employed.

The final goal of the Smart K-Yard is to build the level 4 smart shipyard in the smart shipyard assessment model presented in Table 1. It will be developed with an integrated production system that combines simulation-based engineering system, connected, automated, and intelligent technologies. It has four major technology fields and sub-core technologies to implement the smart shipbuilding process and enhance the smart level.

Table 1 Smart shipyard assessment model

	Production & Planning	**Facility Automation**	**Logistics Automation**	**Factory Operation**	**Supply Chain Management**
Level 5	Integrated intelligent/connected/automated-based process life cycle operation				
		Factory integration automation		DT-based forecasting operations	Web-based DT network collaboration
Level 4	Simulation batch automation	Integrated control automation of production facility	Integrated control automation of logistics facility	Real-time factory control	Web-based collaboration
Level 3	Automated full-process production information	Automated full-process production facilities	Automated full-process logistics facilities	Real-time decision making	Dedicated app-based collaboration
Level 2	Automated part-process information	Automated full-process production facilities	Automated full-process logistics facilities	Individual system operation	Product/production information sharing collaboration
Level 1	Manual operation	Manual operation	Manual operation	Manual operation	Wire/email collaboration

The four major technical fields are: 1) intelligent shipbuilding production design platform, 2) shipbuilding process automation technology, 3) shipyard operational efficiency improvement technology, and 4) simulation-based virtual production platform. Intelligent shipbuilding production design platform technology utilizes the latest ICT to support advanced intelligence in the production design environment, which strengthens the global competitiveness of locally made three-dimensional production design software, and effectively supports the production activities of smart yards to prevent loss of materials, malfunctions, and errors in the design phase.

2.3. Business model of Smart K-Yard

The environment of shipbuilding production is changing rapidly owing to the introduction of the fourth industrial revolution technology and the beginning of an aging society. As a result, shipbuilding automation technology is needed to develop production and logistics automation system technologies that can innovate complex shipbuilding production environments to minimize gaps in expertise and maximize production efficiency.

Today, the Korean shipbuilding production environment, which relies on the experience and knowledge of experts, lacks a proactive system to predict and manage high levels of volatility and uncertainty. In addition, a system that can objectively judge and verify innovative efforts for technology development is lacking.

In the field of shipbuilding automation technology, sub-core technologies consist of an intelligent production system for hull and profile, block assembly process automation, piping spool, and outfitting production system and smart logistics system.

Next, the shipyard operational efficiency technology field aims to reduce various waste elements by improving the current operation system of the shipyard, and to develop technologies that can systematically integrate comprehensive supply network business processes encompassing the shipyard and its business partners through acquiring real-time production information. Figure 2 shows the concept of the Smart K-Yard business model.

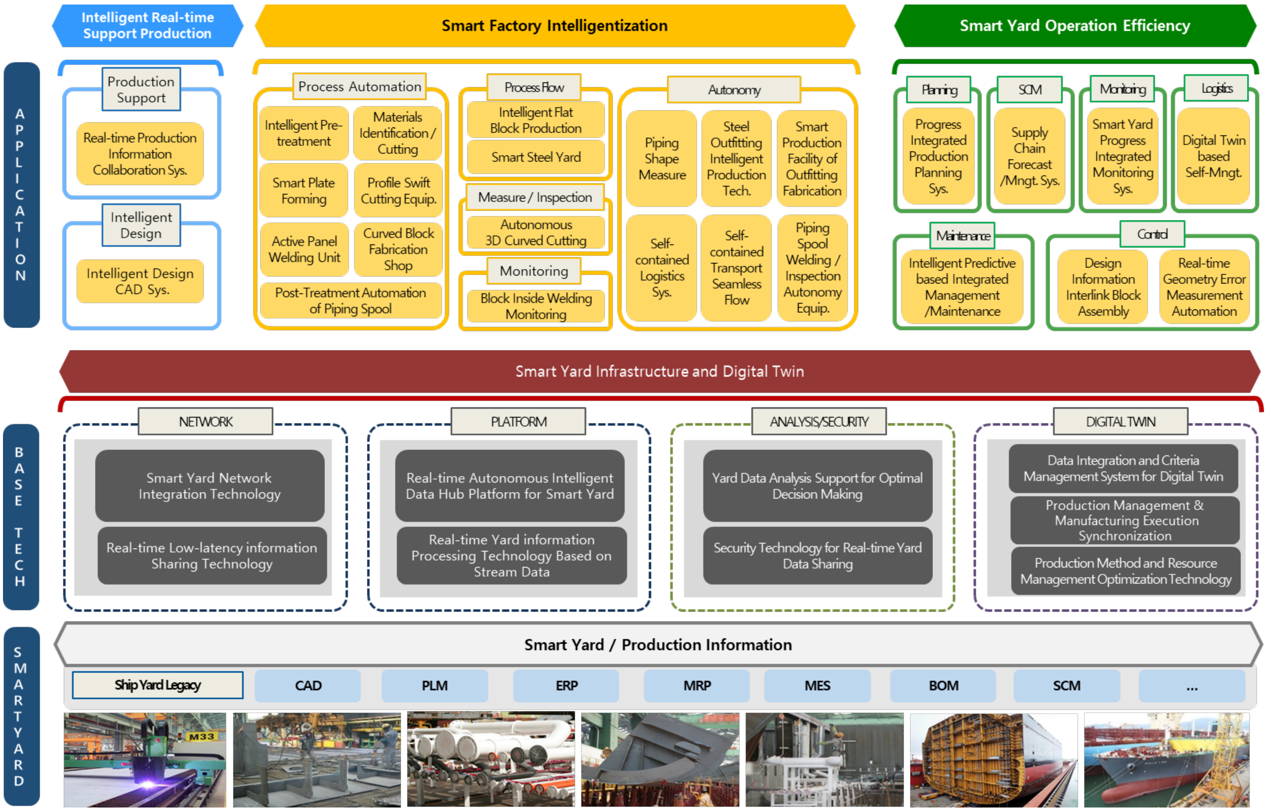

Intelligent Real-time Support Production
Smart Factory Intelligentization
Smart Yard Operation Efficiency
APPLICATION
Production Support
Real-time Production Information Collaboration Sys.
Intelligent Design
Intelligent Design CAD Sys.
Process Automation
Intelligent Pre-treatment
Materials Identification / Cutting
Smart Plate Forming
Profile Swift Cutting Equip.
Active Panel Welding Unit
Curved Block Fabrication Shop
Post-Treatment Automation of Piping Spool
Process Flow
Intelligent Flat Block Production
Smart Steel Yard
Measure / Inspection
Autonomous 3D Curved Cutting
Monitoring
Block Inside Welding Monitoring
Autonomy
Piping Shape Measure
Steel Outfitting Intelligent Production Tech.
Smart Production Facility of Outfitting Fabrication
Self-contained Logistics Sys.
Self-contained Transport Seamless Flow
Piping Spool Welding / Inspection Autonomy Equip.
Planning
Progress Integrated Production Planning Sys.
SCM
Supply Chain Forecast /Mngt. Sys.
Monitoring
Smart Yard Progress Integrated Monitoring Sys.
Logistics
Digital Twin based Self-Mngt.
Maintenance
Intelligent Predictive based Integrated Management /Maintenance
Control
Design Information Interlink Block Assembly
Real-time Geometry Error Measurement Automation
Smart Yard Infrastructure and Digital Twin
BASE TECH
NETWORK
Smart Yard Network Integration Technology
Real-time Low-latency information Sharing Technology
PLATFORM
Real-time Autonomous Intelligent Data Hub Platform for Smart Yard
Real-time Yard information Processing Technology Based on Stream Data
ANALYSIS/SECURITY
Yard Data Analysis Support for Optimal Decision Making
Security Technology for Real-time Yard Data Sharing
DIGITAL TWIN
Data Integration and Criteria Management System for Digital Twin
Production Management & Manufacturing Execution Synchronization
Production Method and Resource Management Optimization Technology
SMARTYARD
Smart Yard / Production Information
Ship Yard Legacy
CAD
PLM
ERP
MRP
MES
BOM
SCM
...
M33

Figure 2: Business Model of Smart K-Yard

2.4. Digital twin-based modelling and simulation platform

Digital twin technology for verification and optimization of shipyard production process, construction method and yard operation is implemented to predict the effect of production, verify efficiency, optimize simulation-based process, validate and verify a new construction method. The digital twin shown in Figure 3 offers a solution to integrate digital models and physical models for diagnosis and prediction of performance, efficiency, and longevity of machinery, equipment, plants, etc. The concept aims to identify the current status through data entered from virtual reality modeling and to improve operation value by responding to changes in real time.

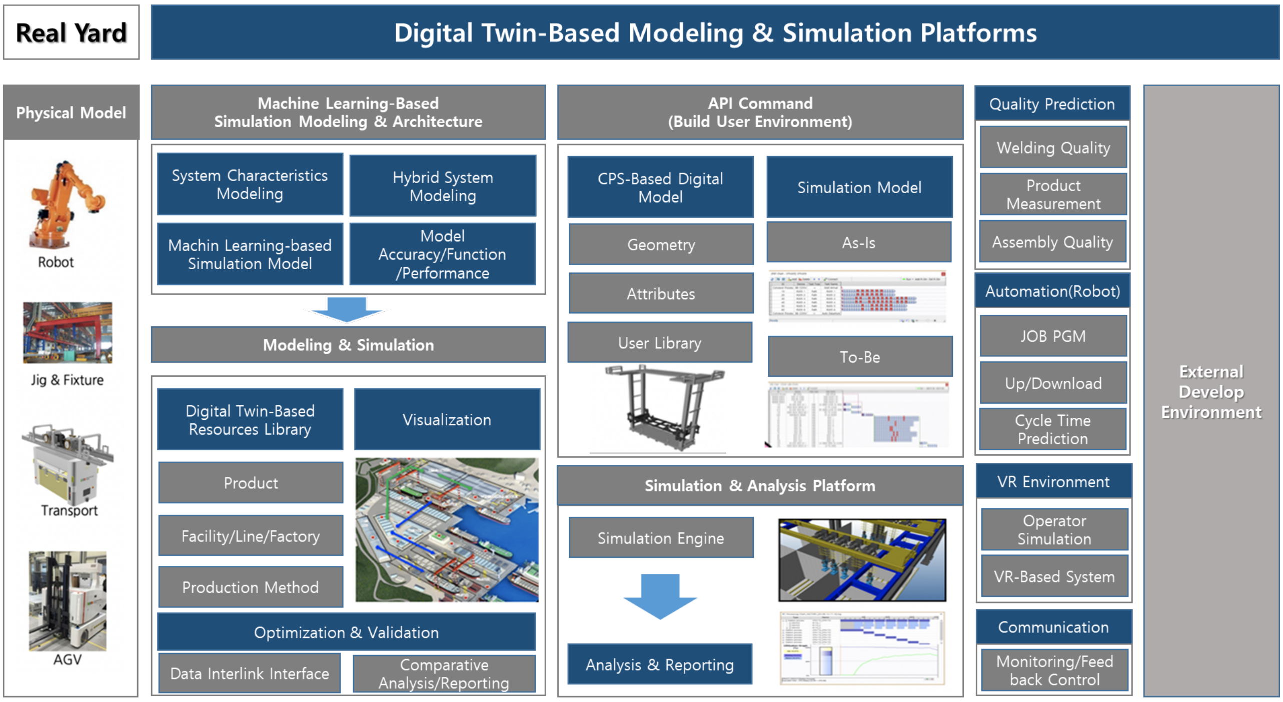


Figure 3: Digital Twin-Based Platforms and Simulation Flatform

Therefore, optimal production and supply chain planning management technologies for pre-production process using cyber-physical production system (CPPS) and IoT, enterprise quality management technology, and real-time integrated control automation and advance intelligence technology, are defined as lower technologies.

Digital mock-up support system: To build a three-dimensional digital mock-up of a shipyard environment such as production equipment (cranes, welding machine, steel cutting equipment), it is possible to identify the assembly conditions and processes in real time in connection with the CPPS platform and to prepare immediate countermeasures in the event of unstable processes.

Control—identify the operation status of a shipyard: Platforms can be employed to build digital twin yards or factories in virtual environments to obtain information about current production methods, facilities, utilities, and workers. In addition, the physical features, status, and properties of production facilities in the yard or plant can be implemented in the digital twin virtual environment and monitored using the product's production information. The current production status can be utilized to detect abnormalities, and formulate corrective response measures in digital twin yards or factories that employ physical production environmental factors and attribute information in virtual environments.

Operation—remote control of shipyard production facilities: It is possible to control the physical production facility by the setting parameters on the digital twin, as well as prevent accidents and achieve energy saving by controlling the crane's emergency stop, lighting or dust air conditioning system by establishing remote control and failure prediction platform.

Optimization—verification of shipyard improvement measures: The effects are verified by applying measures to solve the identified problems and change the production methods in an actual application such as digital twin yards. In addition, a simulation support system is established to optimize production operation by utilizing functional parameters and contextual data of targets.

Figure 4 and Figure 5 show real-time monitoring and failure rate after new application.

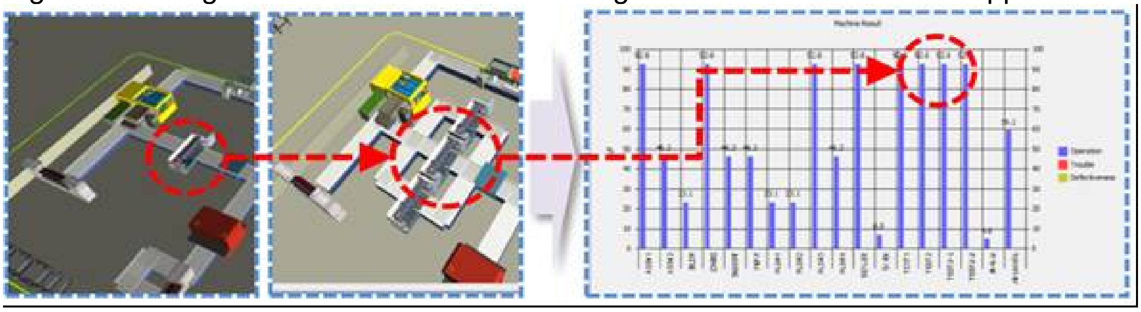


Figure 4: Realtime Monitoring in Facility Layout

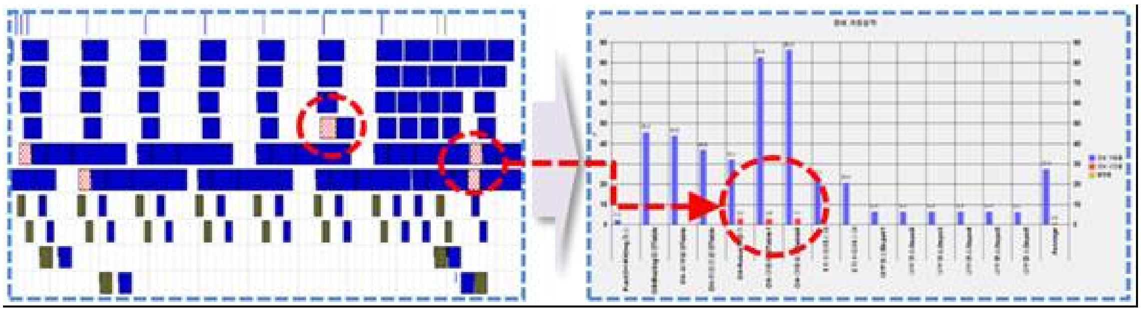


Figure 5: Application of Failure Rate

A set of simulation engineering-based technologies, from the problem diagnosis stage to the development stage and the operational stage, are required to improve productivity, including hardware and software for all the process stages of shipbuilding.

In this project, we defined the field of simulation engineering-based virtualization production platform technology where we intend to create a foundation to eventually build a digital twin yard through the development of shipbuilding CPPS and to sustain the innovation system of the shipbuilding production system.

3. AUTONOMOUS SHIP

3.1. Concept and requirements of autonomous ship

The IMO 98th Maritime Safety Committee has initiated plans for autonomous ship, which was previously recognized only as future technologies; thus, its realization is increasingly becoming a reality. Until recently, autonomous vessels have been referred to using various terms such as smart ship, digital ship, connected ship, remote ship, unmanned ship, and autonomous ship, and are described as Maritime Autonomous Surface Ship (MASS) by the IMO.

Autonomous ship is a comprehensive system to ensure the efficiency of unmanned, autonomous, and transport vessel in a step-by-step upgrade for safety, reliability, and efficiency.

① **Safety**

Autonomous vessels can prevent accidents using appropriate technologies in terms of avoiding collisions with ships and obstacles, steering avoidance in bad weather conditions, and improved visibility and situational awareness to prevent human errors. However, in an emergency or accident, initial failure to respond may lead to increased damage and additional safety threats such as hacking.

② **Reliability**

We plan to improve reliability in ship operation using systematic technical problem-solving skills (centralized maintenance) in autonomous vessels and land, rapid detection and warning, preventive maintenance, and simpler ship hull design, rather than the judgment of a few people on board the ship.

③ Efficiency

It is expected that the operating costs will be reduced through the use of system-wide energy efficiency technologies such as unmanned ships, eco-friendly fuel, increased fuel efficiency, real-time route optimization, and will also be more efficient in maritime affairs.

Autonomous vessels are not limited to fully autonomous vessels. We can set such vessels as remotely operated local vessel → remote controlled unmanned coastal vessel → autonomous unmanned ocean-going ship. Appropriate technologies are needed to verify each phase.

Thus, from level 3 of autonomous driving, autonomous vessels can be an extension of land-based services such as remote control, remote diagnostics, and remote maintenance. This is a step in which ICT provides efficiency and reliability in terms of commercialization. Furthermore, support services will be provided to handle various maritime services on land by reducing or unmanned personnel, and infrastructure that supports safe operation of ships such as buoy and lighthouse will be intelligent and converted.

In addition, connectivity of data is necessary to achieve level 3 of autonomous ship. E-navigation (S-100) can improve the connectivity of data and the communication environment between ship and land (diffusion of VSAT and speed of technology evolution), as well as standardization of in-vessel communication (NMEA, Modbus, ISO DIS19848, etc.).

3.2. Unmanned application of autonomous vehicle technology using artificial intelligence

It will be possible to utilize autonomous technology for autonomous vessels by employing autonomous vehicle application technology. Level 4 or above autonomous cars can be defined as monitoring the driving environment and the fallback function for system errors that allows the system to respond autonomously without driver intervention.

According to the NHTSA autonomous vehicle introduction scenario, the early introduction of an autonomous vehicle into the market, even if it is not perfect, will help reduce fatalities in the long term. It is estimated that the introduction of a self-driving car, which is 10% safe compared with average human driving in 2020, will reduce the number of deaths from traffic accidents by approximately 520,000 compared with the introduction of a fully autonomous vehicle in the market in 2040. Therefore, the above implications of autonomous vehicle should be considered in reviewing the safety conditions of autonomous ships.

Sensor fusion technology for monitoring the driving environment is being developed from the rule-based approach level to deep learning-based level, and it is approaching or exceeding the human perception level. Artificial intelligence can improve the autonomous level through iterative learning on various driving environments; thus, it is essential to acquire driving data for learning. According to the book "Deep Learning" by Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016), there are approximately 5,000 learning data sets per category that yield acceptable performance, and at least one million learning examples are required to match or exceed human performance. In order to accurately achieve learning in artificial intelligence, it is necessary to tag the correct answer for each learning example, which is time consuming.

In addition to object recognition, autonomous navigation artificial intelligence technology requires a variety of context awareness, collision assessment, unexpected response, driving range extraction, and even end-to-end. Currently, artificial intelligence technology is mainly used in the field of cognition, and studies employing a deep learning model for object searching are increasingly conducted in cognitive fields using image sensors.

Artificial intelligence technology was first employed in autonomous driving object recognition, and in decision making based on a complex road situation. However, it is necessary to investigate different application methods for maritime situations in which autonomous ship are applied differently from autonomous vehicles.

Figure 6 shows the structure of the computing module for artificial intelligence of autonomous vehicle. The module is divided into a platform part composed of HW and OS, a deep learning part composed of a deep learning framework and model, and an application part composed of recognition, judgment, and control.

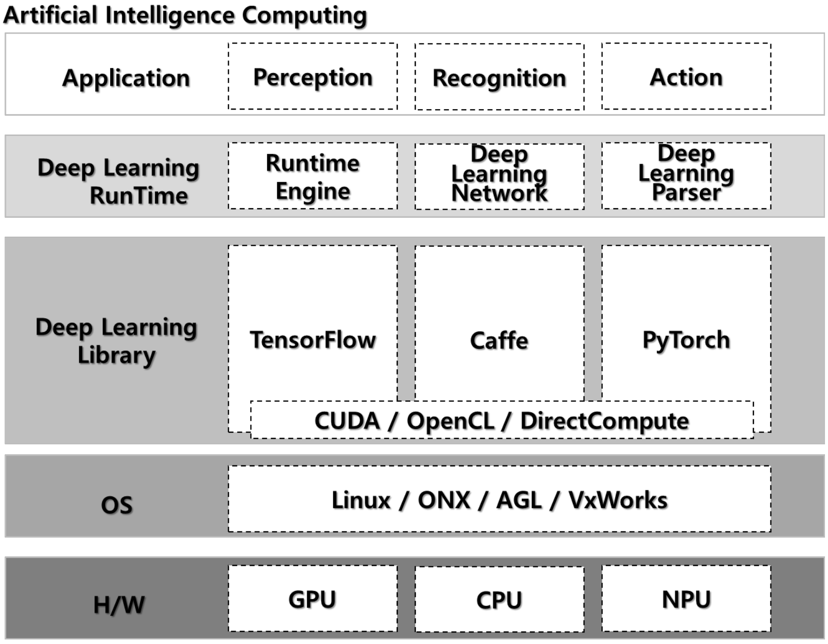


Figure 6: Structure of Computing Module for Autonomous Vehicle

4. CASE STUDY OF AUTONOMOUS SHIP

4.1. Unmanned ship development trend in Korea

Currently, the development of unmanned ships has continued to increase around the world. South Korea is still in the initial stage of research in large unmanned commercial vessels, and technology for unmanned small ships is being developed mainly in laboratories and military facilities.

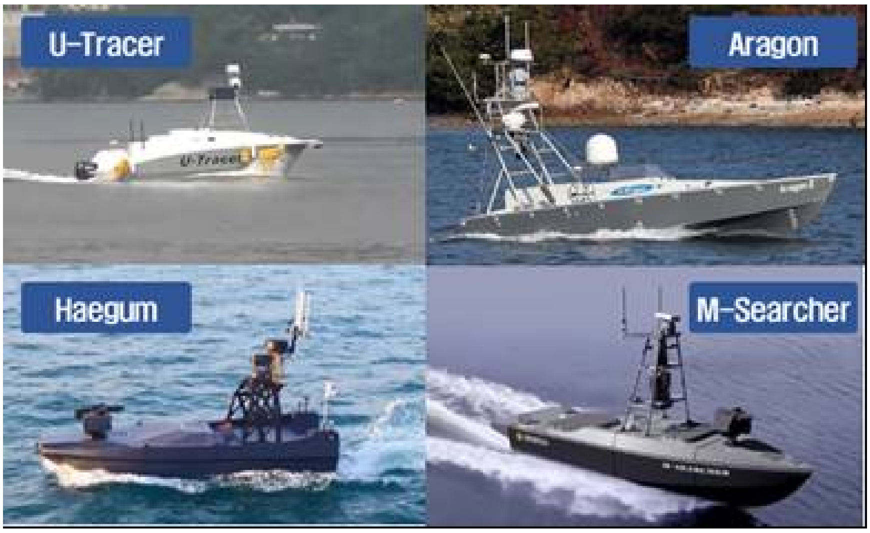


Figure 7: Examples of Unmanned Ship Development in Korea

The U-Tracer in Figure 7 is an outboard type unmanned water reactor developed by the Agency for Defense Development in 2015 and equipped with underwater acoustic target technology that can autonomously track obstacles and targets in water. Aragon is a multipurpose intelligent unmanned ship developed by a private research institute in 2015; it can operate up to 20 kilometers from the ground control system and was developed for marine research and surveillance purposes. Haigeum is an unmanned ship with maritime weapons system developed in 2015 for military operations such as surveillance reconnaissance, mine exploration, and other missions in coastal waters. M-Searcher is an unmanned water reactor developed by the Agency for Defense Development in 2019 to carry out various tasks such as surveillance and underwater search and rescue.

In the future, Korea plans to conduct research on unmanned water vessels for ports/base areas. The technology presented in this paper was used on a platform mounted on the M-Searcher.

4.2. USV platform configuration

Figure 8 shows the composition of the USV platform used in this study. The USV platform is divided into the main control systems, engine control systems, waterjet control systems, and power control systems.

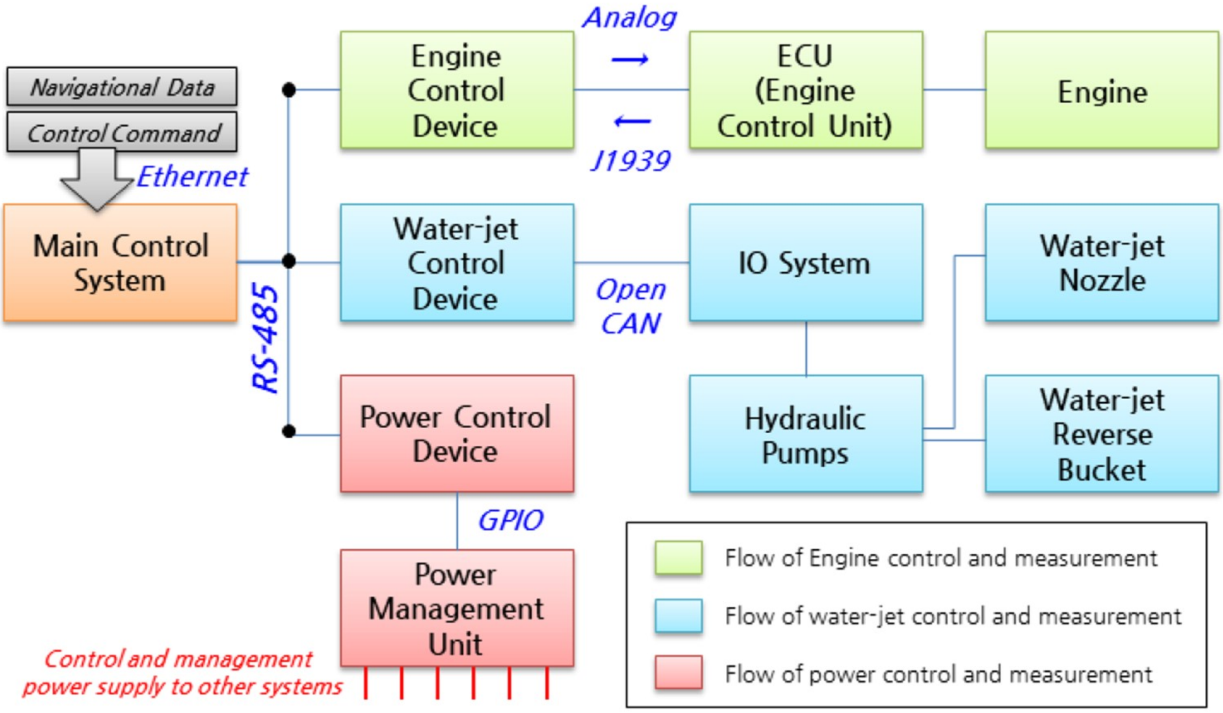


Figure 8: USV Platform Configuration

The main control system receives control commands from the operating system and transmits them to the engine, water jet, and power control systems. It interprets the commands received through Ethernet communication and transmits the corresponding operation to the control device to enable automatic control. Then, the status data of the system received from each device is collected and transmitted at 10 Hz cycle. The engine control unit is responsible for starting, RPM control, and switching functions.

The errors that may occur in this system have not been evaluated as it is still in the development stage. Therefore, the system was configured so that mode switching can be performed for manned control under such condition. In manned mode, the system was designed to ignore commands from the computer, even if such commands were transmitted to engine control. Safety has been improved in the event of an error in the USV platform by allowing human control.

In remote starting, the system was designed to start only when there is no alarm from the ECU status information transmitted through J1983 protocol. The engine status information from the ECU is transmitted to the main control system.

The waterjet control system was also developed to enable the same stable operation as the engine by introducing the manned/unmanned control switchover function. The waterjet controls the nozzles responsible for steering the ship via hydraulic pressure and buckets, which is responsible for forward and backward movements.

In unmanned mode, the system was equipped with a mode function, which is the same as that of the engine, so that it cannot be operated even if an operator arbitrarily manipulates the steering control device. Information (such as the waterjet status and alarms) is transmitted to the waterjet control system at 10 Hz cycle through the I/O system, and the corresponding data is sent to the land control center via the main control system.

The power control unit distributes the power produced by the generator to the platform within the USV. The power supply should be controlled for stable operation of the system in the event of a malfunction because the USV platform is equipped with various devices. The power supply control system monitors the over-current and over-voltage of the system in the ship, automatically shuts down in the event of abnormalities, and sends monitoring data and alarms to the control center in real time.

4.3. Test environment

A test bed with the same engine and water jet was built for testing the USV platform. The engine and water jet used on the platform are the VGT450 from Marin Diesel, Sweden and the AJ285 from Alamari, Finland, as shown in Figure 9.

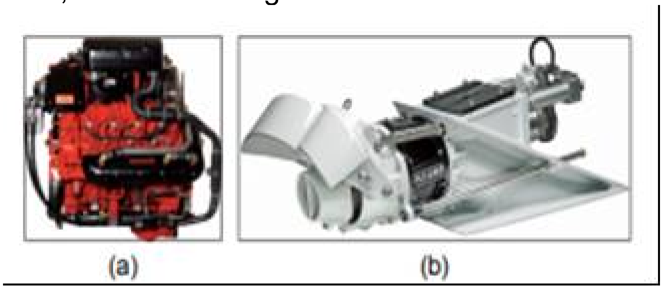


Figure 9: Propulsion equipment (a) Engine and (b) Water-jet

Table 2 Specification of Platform Equipment

Equipment	Specification
Engine (VGT450)	- Max. Power: 450 HP - Max. RPM: 3,600 RPM - Weight 510 kg (Dry)
Water-Jet (AJ285)	- Max. Power: 500 HP - Max. Shaft RPM: 3,700 RPM - Max. Impeller Dia.: 288 mm - Weight: 148 kg
Power (Self-Developed)	- Max. Capacity: approx. 13 kW - Channel: 6 (AC 220 V or DC 28 V) - Weight: 30 kg

The fabricated test bed was fixed on the test bed using a frame suitable for testing in the tank. Figure 10 shows the design of the frame to fix the test bed. Figure 11 shows a frame with the test bed; the test bed was fixed to a water tank, and long continuous operation tests were conducted in the basin.

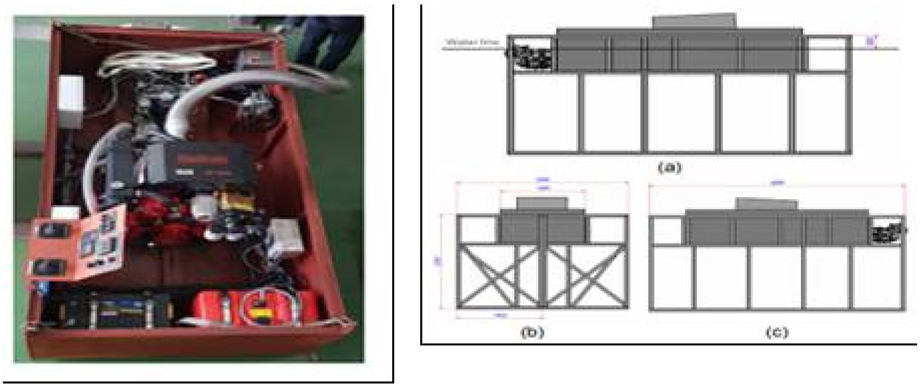


Figure 10: Test Bed for the USV Platform and Frame for the Test Bed

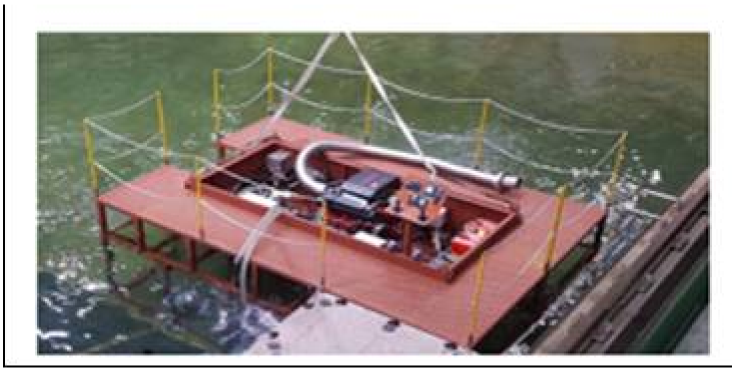
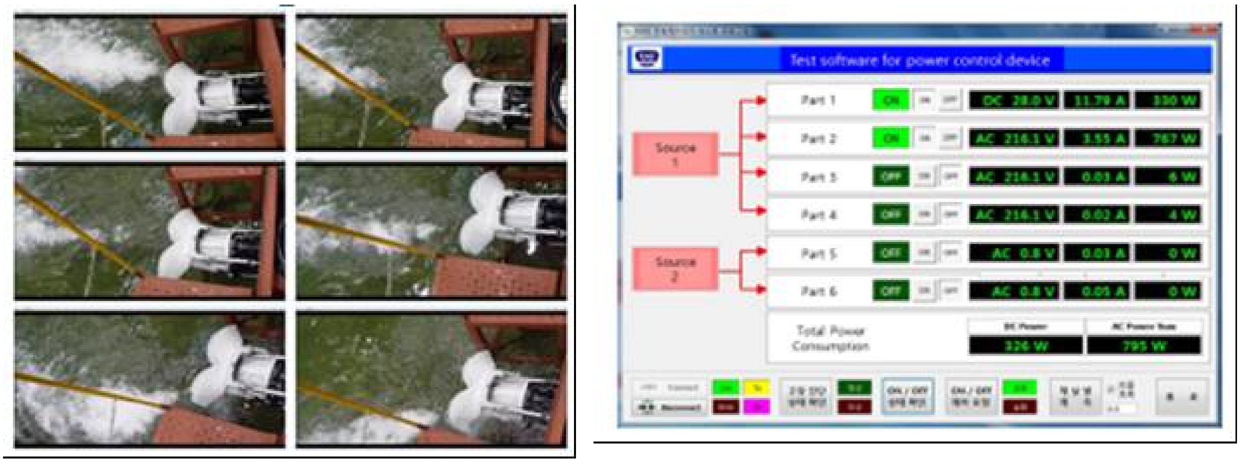


Figure 11: Basin Test of the USV Platform

4.4. Test result

The remote control function of the system was tested using the test bed of the USV platform installed in the basin. Tests were conducted to verify the remote power control, remote engine control, and remote waterjet control functions of the platform.



Test software for power control device

Source 1	Part 1	ON	DC 28.0 V	11.79 A	330 W
	Part 2	ON	AC 216.1 V	3.55 A	767 W
	Part 3	OFF	AC 216.1 V	0.03 A	6 W
	Part 4	OFF	AC 216.1 V	0.02 A	4 W
Source 2	Part 5	OFF	AC 0.8 V	0.03 A	0 W
	Part 6	OFF	AC 0.8 V	0.05 A	0 W

Total Power Consumption: DC Power 326 W, AC Power Sum 795 W

Figure 12: Captured Images from Video during the Test and Test Result of Power Control Device

Figure 12 shows the real-time control performance of the test bed obtained using the program for remote power control. The figure also shows photograph of the remote water jet controlling the engine speed at 1800 RPM, then remotely controlling the steering angle of the water jet in the range of 25° port and 25° starboard; the corresponding motion of the waterjet is also shown. This system is currently installed in the USV platform systems and various tests have been completed.

5. CONCLUSIONS

Despite the decreasing cost competitiveness in the global shipbuilding industry, the market for small and medium-sized ships, such as smart ships and unmanned ships, is still attractive. To ensure that small and medium-sized shipyards, as well as large shipyards, remain competitive, they must shift from the current paradigm of shipbuilding technology and introduce technologies for the development of unmanned ships and the process for shipbuilding.

In this study, we investigated concepts for the development of unmanned ships and autonomous ships, as well as the necessary technology for smart shipbuilding.

Global technological cooperation between international research institutes, universities, and companies is needed for future smart ships in response to the fourth industrial revolution in shipbuilding industries, focusing on IoT sensor technology, robotic system, and unmanned ship technology.

ACKNOWLEDGEMENTS

This work is part of the Marine Technology Convergence and Integration project and unmanned ship of the Agency for Defense Development (ADD, Korea) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and the Ministry of Ocean and Fisheries (MOF, Korea)

REFERENCES

David G. Groves, Nidhi Kalra, "Autonomous Vehicle Safety Scenario Explorer," 2018, NHTSA.

Hun-Gyu Hwang, et al. A Development of Integrated Control System for Platform Equipment of Unmanned surface Vehicle. Journal of the Korea Institute of Information and Communication Engineering, 2017, 21.8: 1611-1618.

HyunWoo Kim, Jangmyung Lee, Robust sliding mode control for a USV water-jet system. International Journal of Naval Architecture and Ocean Engineering, 2019, 11.2: 851-857.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, 2016

Keith Naughton, "Fear of Robot Rides Rises Following High-Profile Road Deaths," 2018, Bloomberg.

Shin-Bae Park, Won-Jae Kim, Kurnchul Lee. A Study of the Development Test and Evaluation and Verification Procedure of a Multi-Mission USV, M-Searcher. Journal of Ocean Engineering and Technology, 2018, 32.5: 402-409.

Yongwon Kwon, Daekyun Oh, Yong-kukJeong, BSNAK,, Vol. 55, No 4, December 2018 "Technical Field and core technologies for smart ship yard and shipbuilding process", P9-P15.

A. Alleyne, R. Liu, A simplified approach to force control for electro-hydraulic systems Contr. Eng. Pract., 8 (12) (Dec. 2000), pp. 1347-1356

Terry Huntsberger, Gail Woodward, Intelligent autonomy for unmanned surface and underwater vehicles OCEANS 2011, IEEE (2011)

Nam-Sun Son, Hyeon-Kyu Yoon, Study on a waypoint tracking algorithm for unmanned surface vehicle (USV) J. Navig. Port Res., 33 (1) (2009), pp. 35-41

J.H. Yoo, J.K. Ryu, Development of Steering System for Autonomous Unmanned Surface Vehicle, vol. 2016, KSPE (2016)

Safety related cyber-attacks identification and assessment for autonomous inland ships

Victor Bolbot[1*], Gerasimos Theotokatos[1], Evangelos Boulougouris[1] and Dracos Vassalos
[1] Maritime Safety Research Centre, University of Strathclyde, UK

ABSTRACT

Recent advances in the maritime industry include the research and development of new sophisticated ships including the autonomous ships. The new autonomy concept though comes at the cost of additional complexity introduced by the number of systems that need to be installed on-board and on-shore, the software intensiveness of the complete system, the involved interactions between the systems, components and humans and the increased connectivity. All the above results in the increased system vulnerability to cyber-attacks, which may lead to unavailability or hazardous behaviour of the critical ship systems. The aim of this study is the identification of the safety related cyber-attacks to the navigation and propulsion systems of an inland autonomous ship as well as the safety enhancement of the ship systems design. For this purpose, the Cyber Preliminary Hazard Analysis method is employed supported by the literature review of the system vulnerabilities and potential cyber-attacks. The Formal Safety Assessment risk matrix is employed for ranking of the hazardous scenarios. The results demonstrate that a number of critical scenarios can arise on the investigated autonomous vessel due to the known vulnerabilities. These can be sufficiently controlled by introducing appropriate modifications of the system design.

Keywords: Safety; Cybersecurity; Autonomous inland vessel; Navigation and propulsion systems; Cyber Preliminary Hazard Analysis.

1 INTRODUCTION

Cyber-Physical Systems (CPSs) represent a class of systems consisting of control elements as well as software and hardware, which are used to effectively control physical processes advancing in a number of application areas including the maritime industry (DNV GL, 2015). CPSs are expected to increase the productivity and safety levels by removing, substituting and/or supporting the operator in the decision-making process, thus reducing the number of human errors leading to accidents. Typical examples of the marine CPSs include the Diesel-Electric Propulsion plant, the Safety Monitoring and Control System, the Dynamic Positioning System as well as the Heating Ventilation Air Conditioning systems (DNV GL, 2015). The number of the CPSs is expected to increase in autonomous ships, which are considered to be the ultimate maritime CPS.

The introduction of the CPSs is accompanied with increased complexity owed to the heterogeneous character of the CPSs, the dependence on information exchanging with other systems, the additional new interactions with humans, the increased number of controllers running complicated software and the increased interconnectivity required for implementing the desired CPSs' functionalities (Bolbot, Theotokatos, Bujorianu, Boulougouris, & Vassalos, 2019). However, this also introduces new hazards as cyber-attacks can exploit vulnerabilities in the communication links and directly affect the integrity or availability of the data and control systems leading the CPSs

---

[*] Corresponding author: victor.bolbot@strath.ac.uk

to accidents (Bolbot et al., 2019; Eloranta & Whitehead, 2016). Considering that ships and their cargo are assets with great value, this inevitably will lead to severe financial consequences in case of an autonomous vessel; it may also have serious safety implications.

There is an increasing number of concerns with respect to the ship systems vulnerability to cyber-attacks in the maritime industry and a number of guidelines have been developed to address these concerns (Boyes & Isbell, 2017; DNV GL, 2016, 2019; IMO, 2016; Maritime affairs directorate of France, 2016; United States Coast Guard, 2015). In addition, a number of previous research studies focused on the cyber security assessment of the ship control systems and ship networks in autonomous ships. Jones, Tam, and Papadaki (2016) reported the identification of different attack scenarios on a cargo ship. Tam and Jones (2019) proposed a model-based approach for the risk assessment of cyber-threats named MaCRA (Maritime Cyber-Risk Assessment) by considering the technological systems vulnerabilities as well as the ease-of-exploit and the potential hackers rewards. Using the same model-based approach, Tam and Jones (2018) implemented a risk assessment for a number of autonomous vessels. Kavallieratos, Katsikas, and Gkioulos (2019) employed the STRIDE (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service and Elevation of Privilege) method to assess risks in an autonomous vessel. Omitola, Downes, Wills, Zwolinski, and Butler (2018) analysed an unmanned surface vessel navigation system using the System-Theoretic Process Analysis for cyber-attacks (STPA-sec) targeting at modifying data that are provided as input to the vessel navigation system.

However, in the previous research studies the risk assessment was implemented considering high level system architecture. Furthermore, the risk assessment in the previous studies identified a number of potential attack scenarios, but did not focus on the safety related consequences. In addition, none of the previous studies conducted a risk assessment of an inland autonomous vessel. Inland autonomous vessel is operating in different environment from the short sea or ocean going vessels, has different system requirements and size and can attract the interest from different hackers groups than the short sea and ocean going vessels.

Therefore, the hazardous scenarios that can arise due to cyber-attacks can be very different in autonomous inland ship. In this respect, the aim of this study is to implement a risk assessment for the navigation and propulsion systems of an inland autonomous vessel. To the best of authors knowledge, this is the first study applying the Cyber Preliminary Hazard Analysis (CPHA) method to an autonomous vessel. The novel contributions of the study include (a) the adjustment of CPHA for application to ship systems, (b) the identification of potential hazardous scenarios arising due to cyber-attacks in propulsion and navigation system of an inland autonomous ship and (c) the highlighting of the critical safety/cyber security control measures for this ship.

The remaining of this paper is organised as follows. The followed method for cyber-attacks risk assessment is presented in Section 2. A description of an inland autonomous vessel navigation and propulsion systems is provided in section 3. In section 4, the results of the method application are provided and discussed. In the conclusions section, the main findings are summarised and suggestions for the future research are provided.

2 METHODOLOGY

During the selection of suitable methods, the following requirements have been considered:

- The method must be aligned with the relevant cyber security standards - IEC 62443, ISO 27000 and IEC 61580, and need to be applicable either during the high-level or the detailed level risk analysis (Flaus, 2019).
- The method must focus on the cyber security induced safety risks (Flaus, 2019).
- The method must incorporate different potential attackers groups (Tam & Jones, 2019).
- The method must be marinised – addressing the needs of maritime industry and aligned with the maritime regulations for safety approval (International Maritime Organisation, 2013).
- The method must be preferentially model-based (Bolbot et al., 2019).

Based on the above considerations the Cyber Preliminary Hazard Analysis (CPHA) (Flaus, 2019) has been selected. The advantages of this method are the following:

- The method can be applied during the initial design stages and does not require many details for the investigated system characteristics (Bolbot et al., 2019) similarly with the STRIDE and MaCRA methods.
- The method is not as labour intensive as STPA (Abdulkhaleq & Wagner, 2015), although it can be less formal approach and less detailed when it comes to hazards identification. Therefore, the CPHA is easier to be applied during high-level risk assessment. The STPA does not have any specific guidance related to identification of cyber attacks, simply suggests that some hazardous scenarios can arise due to cyber security violation (Young & Leveson, 2014). The CPHA also allows ranking of different scenarios which is not integral part of the STPA.
- The method incorporates the available or new safety and security barriers, guiding in this way the system design improvement. This information is not present in the STRIDE and MaCRA methods.
- Compared to the STRIDE and MaCRA methods, the CPHA: (a) is not limited to the specific suggested attack types, and; (b) describes better the relevant hazardous scenarios by incorporating the potential attack type and the relevant hazardous consequences.
- CPHA is based on Preliminary Hazard Analysis (PHA), which is a well-known method for safety assessment and is proposed by ISO 31000 and IEC 61580.

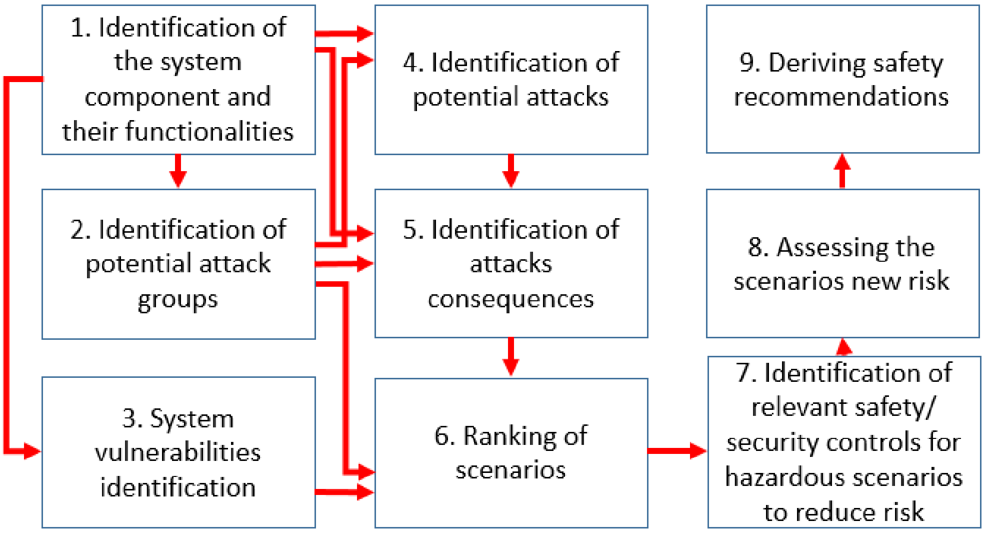

1. Identification of the system component and their functionalities
2. Identification of potential attack groups
3. System vulnerabilities identification
4. Identification of potential attacks
5. Identification of attacks consequences
6. Ranking of scenarios
7. Identification of relevant safety/ security controls for hazardous scenarios to reduce risk
8. Assessing the scenarios new risk
9. Deriving safety recommendations

Figure 1 CPHA methodology flowchart.

The CPHA followed steps are provided in the flowchart depicted in Figure 1, whilst the method steps are elaborated further below. These are the CPHA steps described in (Flaus, 2019) with small modifications. Another difference is that the scenarios ranking is implemented using Formal Safety Assessment risk matrix (International Maritime Organisation, 2013).

The prerequisite for the CPHA is the identification of: (a) the control system elements, (b) the control system elements interfaces with the physical word, the controlled processes and other control system elements interfaces, (c) the potential entry points into system. This is implemented in step 1 (Figure 1), by analysing the available system information as well as by developing the system physical and logical mapping (Flaus, 2019).

As the attackers do not have neither the same motives nor the same resources when attacking a ship network (Tam & Jones, 2019), for identifying and ranking the attack scenarios in step 5 (Figure 1), the following parameters need to be considered: (a) which entry points can be exploited, and; (b) which system will be targeted and (c) in which way by each attacker group. In this respect, the potential attack groups are identified in step 2 (Figure 1) by referring to the relevant literature.

The known vulnerabilities and the potential entry points are identified in step 3 (Figure 1) by using the information provided in the following resources: (a) previous research publications e.g. (Flaus, 2019; Kavallieratos et al., 2019; Omitola et al., 2018; Tam & Jones, 2018); (b) the available maritime standards (Boyes & Isbell, 2017; DNV GL, 2016; IMO, 2016; Maritime affairs directorate of France, 2016); (c) relevant generic standards (IEC, 2011a), and; (d) the Cybersecurity and Infrastructure Security Agency (CISA) database (CISA, 2019a).

The potential vulnerabilities in the system are used to develop the potential attack scenarios in step 4 (Figure 1) (Flaus, 2019). The information about the system interactions and system components functionalities is used to derive the potential consequences in step 5 (Figure 1). In step 6, the scenarios are ranked according to the expected frequency occurrence and the severity of consequences. The frequency and the severity of each attack scenario are ranked using the Formal Safety Assessment (FSA) suggested ranking tables (International Maritime Organisation, 2013), presented in Table 1 and Table 2, whilst the risk is evaluated using the risk matrix presented in Table 3 to harmonise the analysis results with the relevant IMO Formal Safety Assessment guidelines. The frequency ranking for each attack scenario is implemented by considering (a) the level of exposure of each system to attack due to connectivity, (b) the interest of specific attack group in an attack scenario, (c) the attacker level and (d) the access control to the systems. The severity ranking is implemented based on consequences. The preventive and mitigating barriers are identified and proposed in step 7. Then, the scenarios risk is reassessed considering the available or the preventive and mitigating barriers. Based on this analysis results, the relevant safety recommendations at the initial ship design stage are derived. These results can be used as input to more detailed analysis as required by IEC 62443 (BSI, 2009).

Table 1 Ranking for successful attack scenarios (International Maritime Organisation, 2013).

Ranking (FI)	Frequency	Definition	F (per ship year)	F (per ship hour)
7	Frequent	Likely to occur once per month on one ship	10	$1.14\ 10^{-3}$
5	Reasonably probable	Likely to occur once per year in a fleet of 10 ships, i.e. likely to occur a few times during the ship's life	10^{-1}	$1.14\ 10^{-5}$
3	Remote	Likely to occur once per year in a fleet of 1,000 ships, i.e. likely to occur in the total life of several similar ships	10^{-3}	$1.14\ 10^{-7}$
1	Extremely remote	Likely to occur once in the lifetime (20 years) of a world fleet of 5,000 ships.	10^{-5}	$1.14\ 10^{-9}$

Table 2 Ranking for severity of consequences (International Maritime Organisation, 2013).

Ranking (SI)	Severity	Effects on human safety	Effects on ship	Oil spillage definition	S Equivalent fatalities
4	Catastrophic	Multiple fatalities	Total loss	Oil spill size between < 100 - 1000 tonnes	10
3	Severe	Single fatality or multiple severe injuries	Severe damage	Oil spill size between < 10 - 100 tonnes	10^{-0}
2	Significant	Multiple or sever injuries	Non-severe ship damage	Oil spill size between < 1 - 10 tonnes	10^{-1}
1	Minor	Single or minor injuries	Local equipment damage	Oil spill size < 1 tonne	10^{-2}

Table 3 The risk matrix (International Maritime Organisation, 2013)

Risk Index (RI)					
FI	Frequency	Severity (SI)			
		1	2	3	4
		Minor	Significant	Severe	Catastrophic
7	Frequent	(H) 8	(H) 9	(H) 10	(H) 11
6		(M) 7	(H) 8	(H) 9	(H) 10
5	Reasonably probable	(M) 6	(M) 7	(H) 8	(H) 9
4		(M) 5	(M) 6	(M) 7	(H) 8
3	Remote	(L) 4	(M) 5	(M) 6	(M) 7
2		(L) 3	(L) 4	(M) 5	(M) 6
1	Extremely remote	(L) 2	(L) 3	(L) 4	(M) 5
High (H) =Intolerable Risk		Medium (M) =Tolerable Risk		Low (L) =Negligible Risk	

3 CASE STUDY DESCRIPTION

The proposed methodology was applied to an autonomous version of a conventional operational Pallet Shuttle Barge (PSB) (Blue Lines Logistics, 2015) as the particular PSB is going to be retrofitted into an autonomous during AUTOSHIP project. The selected autonomous PSB is supposed to operate from/to the port of Antwerp in Belgium and the interconnected canals. The main ship particulars are provided in Table 4. The focus of the analysis was put on this vessel navigation and propulsion systems, as they are considered the most vulnerable to cyber-attacks (BIMCO, 2018). The equipment that is used for the navigation and the propulsion, as well as the relevant interconnections and interactions between the involved subsystems are schematically shown in Figure 2. The network description was developed based on the information provided in (Boyes & Isbell, 2017; Höyhtyä, Huusko, Kiviranta, Solberg, & Rokka, 2017; Maritime affairs directorate of France, 2016; Schmidt, Fentzahn, Atlason, & Rødseth, 2015; Stefani, 2013) and available drawings for similar ships. The actual network interconnections and equipment may differentiate in the final design of this autonomous PSB. The PSB selected components functionalities description is provided in Table 5. For the present analysis, it was considered that the PSB is in fully autonomous operation, so there is no crew onboard the vessel.

Table 4 PSB particulars.

Type	Catamaran
Length	50 m
Breadth	6.6 m
Maximum Draught	2.2 m
Air draught	5.6 m
Maximum cargo load	300 tonnes
Maximum speed	8.1 knots
Engine output	300 hp
Propulsion type	Diesel-mechanical with azimuth propulsion aft and bow thruster at the bow

Table 5 PSB selected components functionalities description.

Component	Functions
Shore control centre	• Monitoring of physical processes • Navigation control • Control over the ship in emergency/manoeuvring operating modes • Implementation of software updates
Connectivity manager	• Control over information flow between the vessel and the shore control centre
Autonomous ship controller	• Monitoring of the processes safety and alarm generation • Control over ship operating modes (emergency, sailing, autonomous, remotely controlled etc.)
Ship control station	• Interface between crew on board and the vessel, allowing the crew to take control over the navigation systems and engine automation systems
Engine automation system	• Machinery components health monitoring
System Control And Data Acquisition (SCADA) server	• Machinery system sensors measurements and alarms data log
Main engine controller	• Control over engine speed • Engine health status monitoring
Generator controller	• Generator speed control • Generator health status monitoring
Azimuth controller	• Azimuth angle control • Azimuth health monitoring
Bow thruster controller	• Bow thruster speed control
Network cabinet	• Interconnection with other systems
Route planning system	• Selecting the route between departure and arrival point based on the traffic in area
Navigation and collision avoidance system	• Navigating within ports and channels • Position holding • Avoiding collision with other vessels and objects
Situation awareness system	• Picture compilations around the vessel
Electronic Chart Display Information System (ECDIS)	• Detecting position of the ship on the map
Voyage Data Recorder (VDR)	• Principal alarms and sensors measurements recording
Very High Frequency (VHF) radio	• Transmitting messages between vessels
Automatic Identification System (AIS)	• Sending and receiving GPS positions, speed, heading, type of ship, next port and estimated time of arrival to and from surrounding ships
Global Maritime Distress and Safety System (GMDSS)	• Sending and receiving critical safety alerts
RAdio Detection And Ranging (RADAR)	• Detection and determination of the position and speed of the objects
Light Detection And Ranging (LiDAR)/ Laser Detection And Ranging (LADAR)	• Detection and determination of the position and speed of the objects with greater accuracy
Video cameras	• Objects detection and recognition
Echo sounder	• Depth measurement
Global Positioning System (GPS)	• Position measurement, and indirectly speed measurement
Gyro compass	• Angular position and velocity measurement
Speed log measurement	• Speed measurement

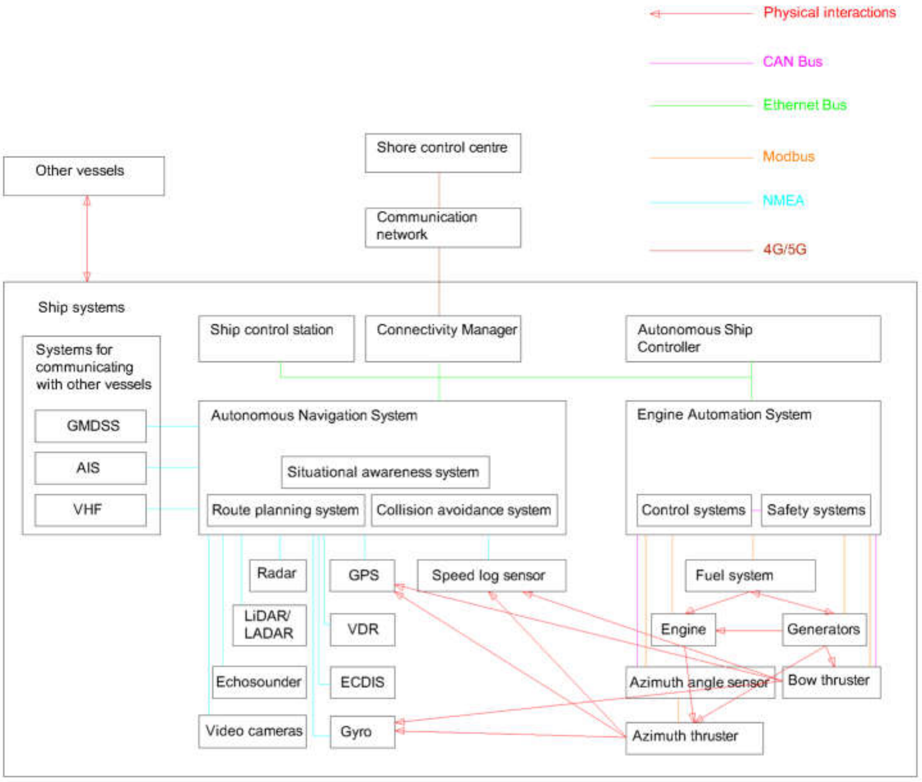


Figure 2 Schematic of PSB network and interactions

4 RESULTS AND DISCUSSION

The investigated autonomous ship systems control elements, their interactions with other control elements, the potential entry points and the relevant network zones are presented in Figure 2 and Figure 3, which are the results of the used methodology first step.

The potential attackers can be classified into the following groups (Results of step 2) (Boyes & Isbell, 2017; Flaus, 2019; IEC, 2011b; Tam & Jones, 2019):

- Former malicious employees aiming at taking revenge from the ship operating company.
- Malicious external providers desiring to steal the machinery data.
- Activists opposed to autonomous ships introduction in the maritime industry (Hacktivists).
- Hackers willing to prove and train their skills.
- Competitors aiming at stealing valuable data or sabotaging and damaging the ship.
- Criminals aiming at stealing the ship, its cargo, components or seeking for a monetary reward.
- Terrorists aiming at damaging the ship and/or causing fatalities.
- States in case of total war aiming at damaging or taking control over the ship.

Since the terrorist group is the group of people targeting the most on the accident achievement, the focus of the present case study will shift towards identifying attacks and safety scenarios, which may be of interest by terrorists. For this analysis, it was assumed that there is an undisclosed group of terrorists which possesses significant technical knowledge about the vessel and its communication systems. This group attacks can be considered similar to the attacks implemented by states in case of a total war. The potential vulnerabilities that can be exploited and the attacks that can be realised are provided in the following paragraphs.

Social engineering attacks are considered the most powerful tool on the hackers hands (Flaus, 2019). Thus, a successful phishing scam can be used to get access of the ship through the shore control centre. Attacks installing malware using flash medium can be also implemented on the shore control centre and at ship control station, as described in (Lund, Hareide, & Jøsok, 2018) or through accidental communication bridges developed between the smart devices with wireless connectivity used by maintenance personnel and the ship control systems (Oates, Roberts, & Twomey, 2017). 4G protocol has been found vulnerable to a number of attacks, where a malicious node can be used to impede the communication or to steal information (Hussain, Chowdhury, Mehnaz, & Bertino, 2018). However the ship satellite communications systems have been also proved to be vulnerable to penetration (Munro, 2017). Configurations in the communication between the ship and the shore control centre including an anonymous File Transfer Protocol can lead to a cyber security breach (IEC, 2011b). Even Virtual Private Networks can have exploitable vulnerabilities, such as the use of outdated communication protocols (DNV GL, 2016; Flaus, 2019). Remote access can be also facilitated by using an available web link to the system equipment with inappropriate username and password (Munro, 2017; Oates et al., 2017) or due to inappropriate remote unit firewall configuration settings (CISA, 2019d; DNV GL, 2016; Oates et al., 2017).

Physical attacks (Flaus, 2019) can be also considered in the case of PBS as the vessel is operating in a close proximity to the shore (or river/channel banks) and no crew is present. The Programming Logic Controllers (PLCs) can be vulnerable to Denial of Service (DoS) or malware attacks due to an unchecked integer overflow vulnerability (Flaus, 2019) or other vulnerabilities (CISA, 2019b; Oates et al., 2017). Considering that patching may not be as frequently implemented as required and that due to the extensive ship lifetime compared to other information technology systems it may not be technically feasible to patch the software (Oates et al., 2017). Therefore it is highly likely that known vulnerability is being exploited (Nazir, Patel, & Patel, 2017; Oates et al., 2017). However system patching by system provider itself opens new opportunities for attacks as it requires remote connection to the vessel and can allow malware propagation from the software owner (Oates et al., 2017). System hardware can be already infected with malware installed before actual installation on the ship (logic bombs and backdoors) which cannot be captured by functional testing (Oates et al., 2017). An attacker can even freeze one sensor measurement in a PLC, misleading in this way the operator (Krotofil et al., 2014). It is even possible to modify the sensor measurement and trigger a faulty safety alarm (Shinohara & Namerikawa, 2017). The navigation computer systems can be infected using SQL injections (DNV GL, 2016; Flaus, 2019) and the ship navigation systems have been proved vulnerable to malware installations (Wingrove, 2018).

GPS signal is a relatively weak signal and can be easily jammed (Borio, Driscoll, & Fortuny, 2012; Boyes & Isbell, 2017; Farid, Ahmad, Ahmed, & Rahim, 2018), spoofed (Goward, 2017) or resent with delay (Omitola et al., 2018) . AIS information is transferred using VHF radio with no encryption allowing valuable information to be easily obtained (Maritime affairs directorate of France, 2016) but it can be also altered or jammed (Balduzzi, Pasta, & Wilhoit, 2014). LiDAR sensors depend on reflection signal, so they can be spoofed if objects with relevant reflective/absorbent surfaces are set in front of them (Brooks, 2016). Cameras can be easily dazzled or spoofed as well (Alguliyev, Imamverdiyev, & Sukhostat, 2018; Brooks, 2016). The components connected to CAN networks are vulnerable to Denial of Service (DoS) attacks, as an artificial control node can be created in the network, shadowing other controllers, sensors and actuators (Bozdal, Samie, & Jennions, 2018; Kang, Song, Jeong, & Kim, 2018). This generates opportunities for attacks if a physical device can be attached to the ship CAN (CISA, 2019c). Modbus protocol is among the oldest protocols, which is not encrypted and a DoS attack can be easily implemented affecting in this way the availability of sensors/actuators (Flaus, 2019).

More vulnerabilities can be found on Cybersecurity and Infrastructure Security Agency (CISA) website (CISA, 2019a) and National Vulnerability Database (NIST, 2019). For the present analysis though, the above list of vulnerabilities can be considered as adequate.

The CPHA scenarios with RI greater or equal with 9 (Steps 4-8 in Figure 1) are provided in Table 6. In total 48 scenarios have been identified, with 19 of them being critical, 24 in a tolerable region and only 5 of them have been initially characterised as negligible. After the incorporation of the available and new safety/cyber security/security barriers, no scenarios were considered as critical, 21 were considered as tolerable and the rest (27) as negligible. The most critical scenarios are related to the access to the ship control station and shore control station, whilst other top critical ones were related either to the GPS signal related attacks or a malware installation on the collision avoidance system and the situation awareness system. In this analysis, single attacks scenarios have been considered. However, more complicated attacks can be implemented, if several single attack scenarios are combined. Their identification is a subject of detailed risk analysis and hence out of the scope of the present research.

The suggested safety cyber security recommendations (step 9 Figure 1) include the following:

- Increasing redundancy in communication between different network zones (Zone 1, Zone 2, Zone 3 and Zone 4).
- Installation of firewalls between each zone (on the conduits).
- Addition of a safety system verifying the safety of the automatic navigation control system actions.
- Sanity checks and filter application for the GPS signals measurements, addition of anti-interference antennas.
- Encryption for the VHF signals.
- Use of kernels on the critical controllers.
- Two or three factors authentication for software updates and patching.
- Installation of an intrusion detection system in each zone.
- Selecting critical health sensor measurements and sending them to the shore control centre at specific intervals.
- Implementing a safe system shutdown, in case of a critical systems loss.
- Interconnecting the main engine with the generator using power take-in/take off systems.
- Plan route verification by the shore control centre

.

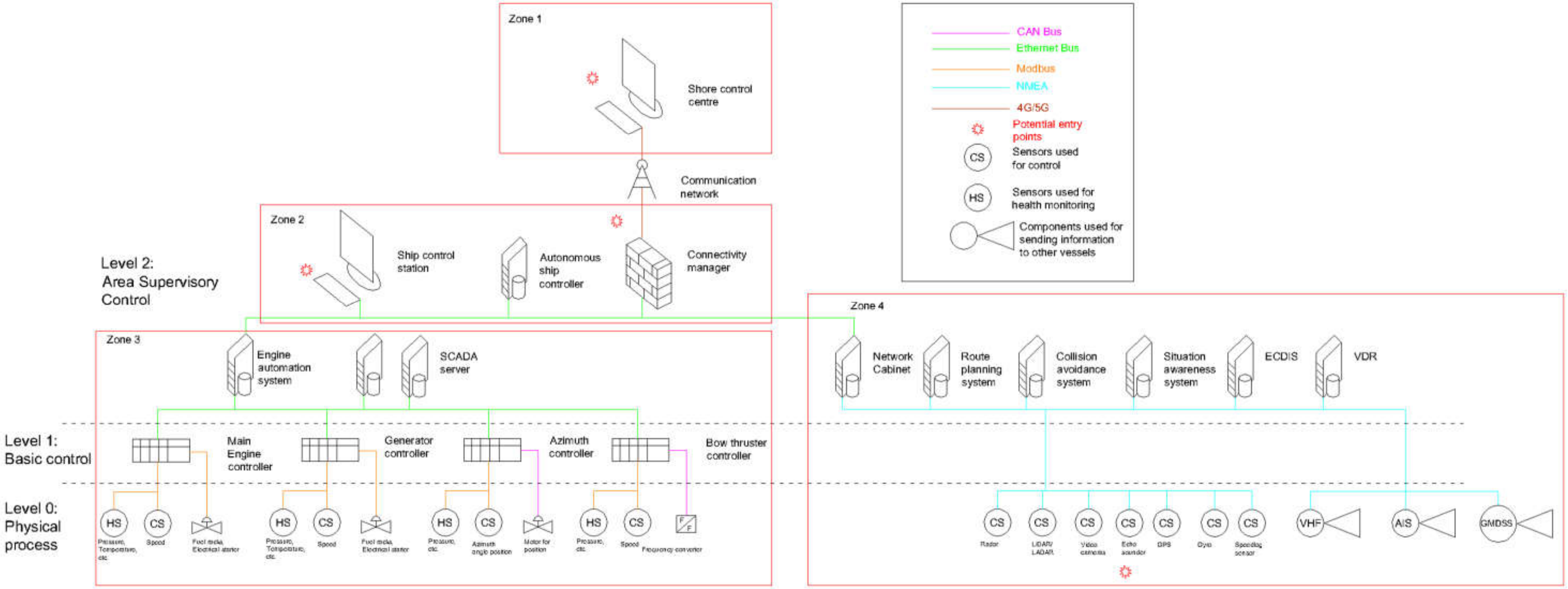


Figure 3 Network logical modelling.

Table 6 The critical CPHA scenarios (initial risk greater or equal than 9).

a/a	System	Attack	Feared event	Consequences	FI	SI	RI=S+SI	Safety/security barriers	S2	SI2	R2=S2+SI2
1	Shore control centre	Social engineering	Stealing access data and gaining authority to perform modifications	Malware installation on ship and loss of ship control	5	4	9	Isolation of shore control centre from the company business network / Closing USB ports / Advanced intrusion detection systems and Antivirus	3	4	7
8	Ship control station	Physical attack	Terrorist in ship control station getting access to the ship control systems	Terrorists gaining control over ship	6	4	10	Two or three factors authentication - Physical barrier to the control room (door, etc.) - Cameras for intrusion detection and alarm - Quick alarm to police – Alarm if cameras are lost	2	4	6
25	Collision avoidance system	Malware installation	System trying to collide with ships or specific objects	Collision/ contact/ grounding	5	4	9	Safety verification system installation / Two or three factors authentication for software modification / Firewall installation / Kernel technologies	3	2	5
27	Situation awareness system	Malware installation	Erroneous picture compilation	Collision/ contact/ grounding	5	4	9	Two or three factors authentication for software modification / Firewall installation / Kernel technologies / Intrusion detection system and Antivirus	2	4	6

5 CONCLUSIONS

The shipping industry is entering new era with autonomous vessels being designed, built and operated. However, their introduction comes at the expense of an increased number of hazardous scenarios due to potential cyber-attacks. In this paper, an enhanced CPHA was employed with the support of the FSA risk matrix for identifying the safety related cyber-attacks, which can be implemented by terrorists, to the navigation and propulsion control systems of an autonomous inland ship.

The main findings of this study are the following:

- A number of technical vulnerabilities such as GPS signal vulnerabilities, PLCs integer overflow vulnerability and VHF lack of cryptography are available at the existing systems, which can be exploited during cyber-attacks.
- Attacks on the shore control centre and the ship control station targeting at getting privileged access have the highest potential safety implications.
- Malware installation on the collision avoidance system and the situation awareness system also have significant safety implications.
- System safety can be improved by adding firewalls on the conduits between different control zones, increased redundancy in communication between control zones and installing intrusion detection systems.

This analysis results can be used to enhance autonomous and other ships designs and guide more detailed risk assessments of the ship systems. The analysis could be extended by applying the CPHA for other attack groups or supporting CPHA results by multiple expert ranking. In addition, a more detailed cyber-security analyses employing more labour intensive methods could be implemented. All this constitute suggestions for future research.

ACKNOWLEDGEMENTS

The work described in this paper was produced in research project AUTOSHIP[†]. The project has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 815012. The authors greatly acknowledge the funding from DNV GL AS and RCCL for the MSRC establishment and operation. The opinions expressed herein are those of the authors and should not be construed to reflect the views of EU, DNV GL AS and RCCL.

REFERENCES

Abdulkhaleq, A., & Wagner, S. (2015). *A controlled experiment for the empirical evaluation of safety analysis techniques for safety-critical software*. Paper presented at the Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering, Nanjing, China. http://delivery.acm.org/10.1145/2750000/2745817/a16-abdulkhaleq.pdf?ip=130.159.52.50&id=2745817&acc=ACTIVE%20SERVICE&key=C2D842D97AC95F7A%2E78A0E221B184F35C%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1517584452_9a89f171925f7a04d4c1fe9971e3675a

Alguliyev, R., Imamverdiyev, Y., & Sukhostat, L. (2018). Cyber-physical systems and their security issues. *Computers in Industry, 100*, 212-223. Retrieved from http://www.sciencedirect.com/science/article/pii/S0166361517304244. doi:https://doi.org/10.1016/j.compind.2018.04.017

Balduzzi, M., Pasta, A., & Wilhoit, K. (2014). *A security evaluation of AIS automated identification system.* Paper presented at the Proceedings of the 30th annual computer security applications conference.

BIMCO. (2018). *Maritime Cyber Survey 2018 - the results*. Retrieved from https://webcache.googleusercontent.com/search?q=cache:rkcjvmcpCakJ:https://www.bimc

---

[†] https://trimis.ec.europa.eu/project/autonomous-shipping-initiative-european-waters

o.org/-/media/bimco/news-and-trends/news/security/cyber-security/2018/fairplay-and-bimco-maritime-cyber-security-survey-2018.ashx+&cd=1&hl=en&ct=clnk&gl=uk

Blue Lines Logistics. (2015). Blue Lines Logistics News. Retrieved from http://www.bluelinelogistics.eu/news

Bolbot, V., Theotokatos, G., Bujorianu, L. M., Boulougouris, E., & Vassalos, D. (2019). Vulnerabilities and safety assurance methods in Cyber-Physical Systems: A comprehensive review. *Reliability Engineering & System Safety, 182*, 179-193. Retrieved from http://www.sciencedirect.com/science/article/pii/S0951832018302709. doi:https://doi.org/10.1016/j.ress.2018.09.004

Borio, D., Driscoll, C. O., & Fortuny, J. (2012, 5-7 Dec. 2012). *GNSS Jammers: Effects and countermeasures.* Paper presented at the 2012 6th ESA Workshop on Satellite Navigation Technologies (Navitec 2012) & European Workshop on GNSS Signals and Signal Processing.

Code of practice - cyber security for ships, (2017).

Bozdal, M., Samie, M., & Jennions, I. (2018). *A Survey on CAN Bus Protocol: Attacks, Challenges, and Potential Solutions.* Paper presented at the 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE).

Brooks, Z. (2016). Hacking driverless vehicles. Retrieved from https://www.defcon.org/images/defcon-21/dc-21-presentations/Zoz/DEFCON-21-Zoz-Hacking-Driverless-Vehicles.pdf

BSI. (2009). Industrial communication networks. Network and system security. IEC TS 62443. In. London, United Kingdom.

CISA. (2019a). CISA - Industrial Control Systems. Retrieved from https://www.us-cert.gov/ics

CISA. (2019b). ICS Alert (ICS-ALERT-17-341-01) - WAGO PFC200. Retrieved from https://www.us-cert.gov/ics/alerts/ICS-ALERT-17-341-01

CISA. (2019c). ICS Alert (ICS-ALERT-19-211-01) - CAN Bus Network Implementation in Avionics. Retrieved from https://www.us-cert.gov/ics/alerts/ics-alert-19-211-01

CISA. (2019d). ICS Alert (ICS-ALERT-19-225-01) - Mitsubishi Electric smartRTU and INEA ME-RTU. Retrieved from https://www.us-cert.gov/ics/alerts/ics-alert-19-225-01

DNV GL. (2015). Technology outlook 2025. In.

DNV GL. (2016). DNVGL-RP-0496 - Cyber security resilience management In.

DNV GL. (2019). Part 6 Additional class notations Chapter 5 Equipment and design features Section 21 Cyber security. In D. GL (Ed.), *Part 6 Chapter 5 Section 21*.

Eloranta, S., & Whitehead, A. (2016). *Safety aspects of autonomous ships*. Paper presented at the 6th International Maritime Conference, Germany, Hamburg.

Farid, M. A., Ahmad, M., Ahmed, S., & Rahim, S. S. (2018). Impact and detection of GPS jammers and countermeasures against jamming. *International Journal of Scientific & Engineering Research, 9*(12), 47-54.

Flaus, J.-M. (2019). *Cybersecurity of industrial systems*. London, United Kingdom: ISTE Ltd.

Goward, A. (2017). Mass GPS Spoofing Attack in Black Sea? Retrieved from https://www.maritime-executive.com/editorials/mass-gps-spoofing-attack-in-black-sea

Höyhtyä, M., Huusko, J., Kiviranta, M., Solberg, K., & Rokka, J. (2017). *Connectivity for autonomous ships: Architecture, use cases, and research challenges.* Paper presented at the 2017 International Conference on Information and Communication Technology Convergence (ICTC).

Hussain, S., Chowdhury, O., Mehnaz, S., & Bertino, E. (2018). *LTEInspector: A systematic approach for adversarial testing of 4G LTE.* Paper presented at the Network and Distributed Systems Security (NDSS) Symposium 2018.

IEC. (2011a). IEC 27005 - Information technology - security techniques - Information security risk management. In.

IEC. (2011b). Information technology — Security techniques — Information security risk management - ISO 27005. In. Switzerland: International Standard organisation.

IMO. (2016). Interim guidelines on maritime cyber risk management. In *MSC.1-CIRC.1526* (pp. 6).

International Maritime Organisation. (2013). *Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process*. London Retrieved from http://research.dnv.com/skj/IMO/MSC-MEPC%202_Circ%2012%20FSA%20Guidelines%20Rev%20III.pdf

Jones, K. D., Tam, K., & Papadaki, M. (2016). Threats and impacts in maritime cyber security.

Kang, T. U., Song, H. M., Jeong, S., & Kim, H. K. (2018). *Automated Reverse Engineering and Attack for CAN Using OBD-II.* Paper presented at the 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall).

Kavallieratos, G., Katsikas, S., & Gkioulos, V. (2019, 2019//). *Cyber-Attacks Against the Autonomous Ship.* Paper presented at the Computer Security, Cham.

Krotofil, M., C, A. A., #225, rdenas, Manning, B., & Larsen, J. (2014). *CPS: driving cyber-physical systems to unsafe operating conditions by timing DoS attacks on sensor signals*. Paper presented at the Proceedings of the 30th Annual Computer Security Applications Conference, New Orleans, Louisiana, USA. http://delivery.acm.org/10.1145/2670000/2664290/p146-krotofil.pdf?ip=130.159.52.50&id=2664290&acc=ACTIVE%20SERVICE&key=C2D842D97AC95F7A%2E78A0E221B184F35C%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1517585100_9ccf2acc8cc4da601c49f53468c4c434

Lund, M. S., Hareide, O. S., & Jøsok, Ø. (2018). An attack on an integrated navigation system.

Cyber security Assessment and protection of ships, (2016).

Munro, K. (2017). OSINT from ship satcoms. Retrieved from https://www.pentestpartners.com/security-blog/osint-from-ship-satcoms/

Nazir, S., Patel, S., & Patel, D. (2017). Assessing and augmenting SCADA cyber security: A survey of techniques. *Computers & Security, 70*, 436-454. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167404817301293. doi:https://doi.org/10.1016/j.cose.2017.06.010

NIST. (2019). National vulnerability database. Retrieved from https://nvd.nist.gov/vuln

Oates, R., Roberts, J., & Twomey, B. (2017). *Chains, links and lifetime: Robust security for autonomous maritime systems*. Paper presented at the Marine Electrical and Control Systems Safety, Glasgow, United Kingdom.

Omitola, T., Downes, J., Wills, G., Zwolinski, M., & Butler, M. (2018). Securing navigation of unmanned maritime systems.

Schmidt, M., Fentzahn, E., Atlason, G. F., & Rødseth, H. (2015). *D8.7: Final report: Autonomous engine room*. Retrieved from

Shinohara, T., & Namerikawa, T. (2017). On the vulnerabilities due to manipulative zero-stealthy attacks in cyber-physical systems. *SICE Journal of Control, Measurement, and System Integration, 10*(6), 563-570.

Stefani, A. (2013). *An introduction to ship automation and control systems*. United Kingdom, London: Institute of Marine Engineering, Science & Technology.

Tam, K., & Jones, K. (2018). *Cyber-risk assessment for autonomous ships.* Paper presented at the 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security).

Tam, K., & Jones, K. (2019). MaCRA: a model-based framework for maritime cyber-risk assessment. *WMU Journal of Maritime Affairs, 18*(1), 129-163. Retrieved from https://doi.org/10.1007/s13437-019-00162-2. doi:10.1007/s13437-019-00162-2

Cyber strategy, (2015).

Wingrove, M. (2018). 'Impregnable' radar breached in simulated cyber attack. Retrieved from https://www.rivieramm.com/news-content-hub/impregnable-radar-breached-in-simulated-cyber-attack-25158

Young, W., & Leveson, N. G. (2014). An integrated approach to safety and security based on systems theory. *Communnications of the ACM, 57*(2), 31-35. Retrieved from http://delivery.acm.org/10.1145/2560000/2556938/p31-young.pdf?ip=130.159.52.50&id=2556938&acc=ACTIVE%20SERVICE&key=C2D842D97AC95F7A%2E78A0E221B184F35C%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1517585608_c8f755bbec0c78ad166f12fbd04cf307. doi:10.1145/2556938

APPENDIX A – ABBREVIATION LIST

AIS	Automatic Identification System
CISA	Cybersecurity and Infrastructure Security Agency
CPHA	Cyber Preliminary Hazard Analysis
DoS	Denial of Service
ECDIS	Electronic Chart Display Information System
FSA	Formal Safety Assessment
GMDSS	Global Maritime Distress and Safety System
GPS	Global Positioning System
LADAR	Laser Detection And Ranging
LiDAR	Light Detection And Ranging
MaCRA	Maritime Cyber-Risk Assessment
PHA	Preliminary Hazard Analysis
PLC	Programmable Logic Controller
PSB	Pallet Shuttle Barge
RADAR	RAdio Detection And Ranging
SCADA	System Control And Data Acquisition
VDR	Voyage Data Recorder
VHF	Very High Frequency

International Seminar on Safety and Security of Autonomous Vessels
17 - 18 September 2019, Helsinki

Prediction Model of Human Error Probability in Autonomous Cargo Ships

Zhang Di^{1,2*}, Zhang Mingyang^{3}, Yao Houjie^{1,2}, Zhang Kai^{1,2}, Fan Cunlong^{1,2}
1．Intelligent Transportation Systems Research Centre, Wuhan University of Technology, Wuhan, China;
2．National Engineering Research Centre for Water Transport Safety, Wuhan, China;
3．Department of Mechanical Engineering, School of Engineering, Aalto University, Espoo, Finland
(Y125 mailbox, No 1178 Heping Avenue, Wuhan University of Technology, Wuhan, Hubei 430063, P.R. China; Tel: +86-139-8616-0037; Fax:+ 86 027-86582280; Email: zhangdi@whut.edu.cn)

ABSTRACT

Despite the use of automation technology in the maritime industry, human errors are still the typical navigational risk factors in Maritime Autonomous Surface Ships with the third degree of autonomy, as defined by the International Maritime Organization. To analyse these human errors, a prediction model for human errors in the emergency disposal process is present. First, the risk factors are identified by analysing the emergency disposal behaviour process of a Shore Control Centre (SCC) under remote navigation mode. This is followed by the establishment of an event tree model of human errors using Technique for Human Error Rate Prediction (THERP). Furthers, a Bayesian Networks (BNs) model based on the THERP is proposed for the three stages: perception, decision, and execution. Subsequently, expert judgments based on the fuzzy theory are used to obtain the basic probability of root nodes and determine the conditional probability of each node in the BNs. Finally, the probabilities of human errors are calculated for the three stages, while the importance of human error factors is quantified with sensitivity analysis, which can provide flexible references for theoretical construction of the SCC and training of staff.

Keywords: Maritime Autonomous Surface Ships; human failure probability; THERP; Bayesian Networks; Risk Control Options

1 INTRODUCTION

With the rapid development of smart ship technology, autonomous ships will inevitably become the main emphasis of innovations by the shipping industry in the near future. For instance, the International Maritime Organization (IMO) at the 98th MSC put the concept of Maritime Autonomous Surface Ship (MASS) forward in 2017. Subsequently, the relevant departments started working on defining applicable laws and regulations from 2018 onwards. Meanwhile, many researchers have come up with preliminary definitions of MASS and established several stages for the development of MASS[1]. Specifically, there can be four development stages based on the perspective of autonomy [2]:

- An automated program can operate ships and provide decision support.
- Ships can be controlled remotely with crew on board.
- Ships can be controlled remotely without any crew on board.
- Ships can be controlled completely autonomously.

The improvement in automation technology will lead to a reduction in the number of people on board, which can promote the realization of autonomous ship navigation[3]. Depending on the extent of development in academic communities that the elaboration of MASS with the third degree of autonomy has already settled down. Examples of projects focusing on MASS with the third degree of autonomy include the *Maritime Unmanned Ships through Intelligence in Networks (MUNIN, 2015)* and *Advanced Autonomous Waterborne Applications (AAWA, 2016)* projects. Specifically speaking, the navigation mode of MASS with the third degree of autonomy can be divided into four subclasses: 1) ships departing from the harbor manually; 2) fully autonomous navigation mode; 3) remotely manipulated driving by officers of Shore Control Centers (SCCs) and 4) fail-to-safe mode [4].

This paper proposes a prediction model of human failure probability with the focus on autonomous MASS with the third degree of autonomy. Maritime risk analysis considers as the research hotspot for both traditional and autonomous ships. On one hand, human errors are the main causes of ship accidents in traditional ships [5, 6], some researchers have argued that the marine safety level could be significantly improved if no crew on board to operate the ship, similar to the MASS with the third degree of autonomy. On the other hand, if there is no crew operated the ship in real-time, new hazard scenarios can emerge as the crew's presence, mobility and flexibility in maintenance and emergency occasions is of the essence. For example, serious accidents are highly likely to occur if there is no crew on board in the presence of equipment failure, such as "shift of cargo" or breakdown of the main engine. Conversely, these problems can be avoided if the crew detects these issues and resolves them on time [7]. Under these circumstances, it is necessary to analyse the role of human errors in the MASS with the third degree of autonomy as the autonomous cargo ships still involve safety risks attached to their operations.

Several studies have conducted human error analysis of maritime accidents. For instance, Ramos et al. [8] discussed the performance of factors influencing human behaviour in autonomous ships. Based on the human cognitive reliability analysis method, these study first subdivided human factors in autonomous ships into direct/indirect and internal/external factors, and subsequently established a decision-making model of factors influencing human behaviour. The authors discussed the main factors that influence the operators' decisions and actions while working on shore were pointed out, four factors: information overload, situation awareness, skill degradation, and boredom in particular. Trudi et al. [9] dubbed the autonomous ship as an "uninhabited" vehicle and argued that it could not be operated without human operators. From a theoretical perspective, autonomous ships can increase the safety of ships. Nevertheless, in reality, there are many uncertainties about safety of the autonomous ships due to lack of first-hand multi-sensory experience. After that, human errors will be transferred to a SCC. Thus, to overcome new challenges that autonomous cargo ships face regarding both safe operation and monitoring, several safety features were put forward, which included communication costs, cyber security, information overload, data sharing, human-machine interaction, situational awareness, psychological load, over-reliance on automatic systems, social factors relating to autonomous cargo ships, and requirements for learning new skills. In addition, Porathe et al. [10] analysed the current situation of SCC for unmanned ships under remote driving mode, and discussed the risks emanating from both ship conditions and human factors that will be faced by the SCC in the near future. The authors observed that maintaining situational awareness in the SCC is much more challenging than creating it. Rather than solely relying on simulated ship bridges, extensive training was needed to maintain situational awareness for real ships. Wróbel et al. [11] assessed the potential impact of unmanned vessels on maritime transportation safety. The human factor issues were focused on remote monitoring and controlling of autonomous unmanned vessels[12,13]. Previous studies mostly focused on human factors from a macro perspective and lacked human error modelling that can occur in the emergency disposal process under remote control of the SCC. In this regard, consolidation is urgently needed of relevant theoretical models to analyse human error factors in autonomous cargo ships[14].

Bayesian Network modelling is an artificial intelligence tool used to model uncertainty in a domain or system with the ability to conduct statistical inference [15, 16], the ability to incorporate new observations into the network, the ability to describe inherent causal and associated probabilistic for the systems and the ability to analyse the complex dependences among the systematic indicators [17]. In the context of human error probability estimation that combines Bayesian approach with an existing method. Some research estimate the human error probability in oil tanker collision [18], winter navigation [19], grounding and collision [20].

This paper aims to fill the above mentioned research gaps and combine the Technique for Human Error Rate Prediction (THERP) and Bayesian Network (BNs) to model the emergency response process followed by operators present on the SCC. Specifically, the THERP prediction method is used to analyse the emergency response process and establish the event tree model. Each event node is modelled with the BNs model, which considers uncertainties and predicts the human error probability.

2 ANALYSIS OF HUMAN EMERGENCY OPERATING BEHAVIOR IN THE SCC

Under the autonomous navigation mode, autonomous cargo ships can be faced with unfavourable situations caused by factors such as external environment, organizational elements, and ship equipment. As these situations cannot be handled on board the ship, danger warnings will be sent to the SCC to seek for assistance from the remote control. In this situation, many researchers have established several decision-making models and control flowcharts corresponding to personnel emergency response process [21, 22]. In our context, the personnel emergency response process of the SCC can be viewed as the process based on the operators' cognitive behaviour, namely, risk information perception → judgment decision → execution [23]. During this process, numerous types of human errors can lead to severe accidents due to influence of the simulation device, equipment, surrounding environment, operating equipment, and personnel quality. This section analyses and establishes the operator's cognitive emergency response process, as shown in Fig. 1.

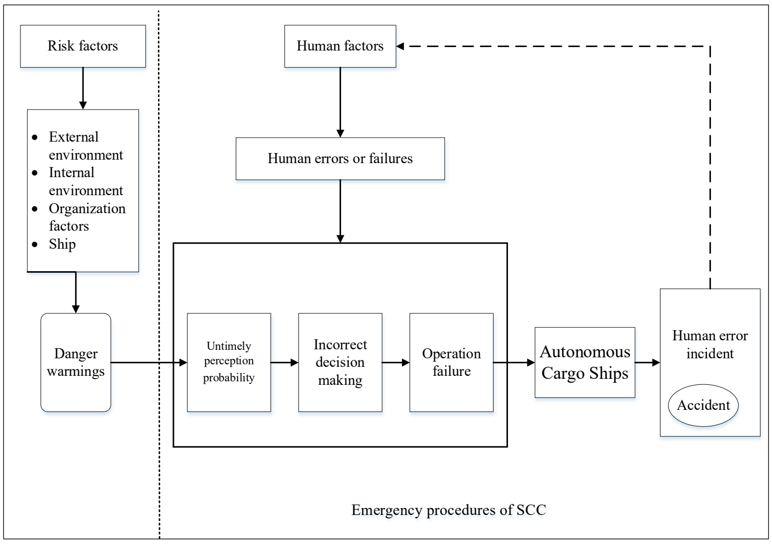


Figure 1. Human error analysis framework on the SCC

3 HUMAN FAILURE PROBABILITY PREDICTION MODEL

3.1 Technique for Human Error Rate Prediction

The THERP can be applied for analysing daily operations following normal regulars, which is widely used in the quantitative analysis of human reliability, complex systems analysis of routine testing and analysis of maintenance tasks.

The THERP involves several aspects such as event tree analysis (ETA) [24, 25], factor analysis of personnel performance, and combining quantitative calculation based on human error database. For instance, based on the event tree model, the THERP creates two attitude branch trees for the time sequence of participating events to calculate the error probabilities of all human behaviours. For this purpose, the THERP needs to consider all kinds of human behaviours in the process of event development, and make accurate quantification based on the specific error characteristics of different operations. This requires the human errors analysis to conduct detailed investigations and interviews on each specific human factor event, to fully understand and identify the key behaviours and related operational details. Thereafter, quantitative analysis can be finally carried out.

Meanwhile, the engineering application is faced with several problems associated with a complicated analysis processes, for example, huge manpower and material inputs, insufficient standardization, and excessive reliance on expert judgment. To overcome these problems, this paper divides the human error process into three stages based on human error rate prediction analysis method, and then constructs the BN model for each stage. These steps mitigate the problem of large amount of data required in the model, and settle down the relationship between human factors greatly [26].

3.2 Bayesian network

The Bayesian formula given in (1) serves as the theoretical basis of the BN. It is principally used to describe the conditional probability inference between two variables.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{1}$$

The formula is made up of prior probability, conditional probability and posterior probability of the events. The prior probability means the occurrence probability of an event based on historical data or subjective expert judgment, conditional probability refers to the occurrence probability of random event B when event A has occurred, under the hypothesis that B is a non-zero probability event. The posterior probability refers to the updated probability of an event occurring after taking into consideration prior and conditional probabilities.

$$P(B|A) = \frac{P(AB)}{P(A)}. \tag{2}$$

The BN is mostly used to model system uncertainties, which are mainly embodied in a Bayesian inference problem. The Bayesian inference problem is a conditional probability reasoning problem, which can be subdivided into two different reasoning models: forwarding reasoning and backward reasoning. Forward reasoning can be viewed as one type of predictive reasoning. To be specific, it transmits the new explanatory variable information forward to the response variable along the direction of the BN arc, thereby updating the probability of the response variable. On the other hand, backward reasoning, also known as diagnostic reasoning, first determines the expected value of the response variable. Then, it places this value in the BN and reverse transmits the information to establish the value of the explanatory variable.

When a BN contains n nodes, it is usually represented as $\Delta=\{G(V, E), P\}$, where $G(V, E)$ represents an acyclic directed graph G containing n nodes. The node variables in the BN graph are represented by the elements in the set $V = \{V_1,...,V_n\}$, the Bayesian arc E stands for the causal relationship between the variables, and P shows the Conditional Probability Tables (CPTs) of nodes in the BN model.

Suppose that an event $\theta = \{\theta_1,...,\theta_n\}$ has n reference values. When the observed values $X=\{X_1, \ldots, X_n\}$ are available, we can calculate the posterior probability distribution table of θ using (3) as follows, based on the BN:

$$P(\theta|x_1, \ldots, x_n) = \frac{P(x_1, \ldots, x_n|\theta)P(\theta)}{P(x_1, \ldots, x_n)}. \tag{3}$$

Figure 2 is an example route from event A to event B in a BNs. Node A impacts node B directly in the network, which means that the former node, being the parent of node B, will affect the occurrence probability of event B. The arrow in Fig. 2 means node A in the directed acyclic graph points to the directed arc of node B, which embodies a sub-node relationship between the two events, while conditional probability P(A|B) represents the dependency between events A and B. Noticeably, while the BNs model is constructed, each node can establish a sub-node relationship with the other nodes, but there should be no circular directed model. That is, closed loop is prohibited for the model.

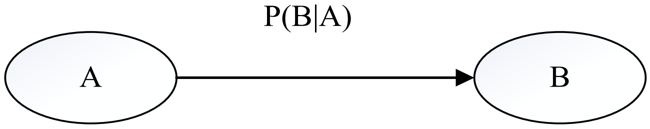


Figure 2. Graphical representation of the basic elements in BNs

3.3 Modeling and Analysis of THERP+BNs

3.3.1 ET model for dealing with emergency

By analysing Fig. 1 and based on the time sequence of events, we can divide the human error events in the SCC into three stages: perception, decision and execution. Accordingly, an event tree model can be established as shown in Fig. 3.

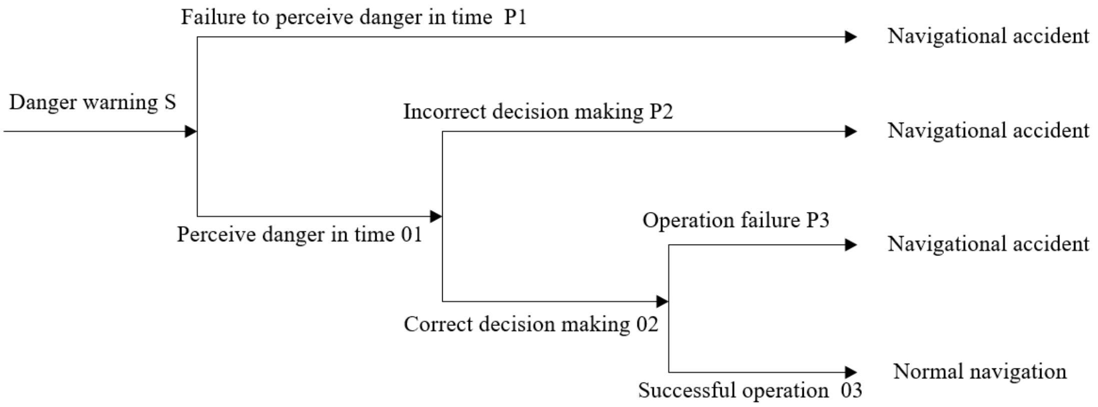


Figure 3. Event tree model for SCC

According to the emergency process to be followed by the personnel during accidents, the human error probability in the event tree consists of the following three parts:

1) Untimely perception probability *P1*, which means that the danger warning is not perceived within the controllable time, and consequently, the control of autonomous cargo ship is not taken over by the SCC in a timely manner.

2) Incorrect decision probability *P2*, which refers to the failure of taking effective measures in an emergency to stop the accident.

3) Operation failure probability *P3*, which means that a correct decision taken by the personnel still leads to an accident.

Therefore, the total human error probability *p* can be obtained as follows:

$$p = 1 - (1 - p_1)(1 - p_2)(1 - p_3). \tag{4}$$

3.3.2 Three-stage human factors classification

The variables of the BNs model are mainly reflected in the form of various nodes in the network. Additionally, the directed edges represent the mutual relationship between these variables, while the conditional probability of the node refers to the strength or degree of dependence of relationships among the nodes. This study strives to establish a Bayesian model of human error for the SCC based on the following steps:

(1) Determination of BNs nodes

In this paper, the human error model of the entire SCC is subdivided into three parts, which include untimely perception, incorrect decision and operation failure. These parts have been used as the output nodes of the three BNs models, respectively. We have identified 16 common human factors that can cause ship accidents, based on literature review and expert investigation of human factors in autonomous ships. In this regard, these 16 factors work as sub-nodes of the BNs and classify them in accordance with the three stages of perception, decision, and execution. To better illustrate the developmental sequence involved in the accident chain of autonomous cargo ships, the classification of these human factors are presented in Table 1.

Table 1. Human factors leading to autonomous cargo ship accidents

Perception stage	Decision stage	Execution stage
A1: Negligence when one person monitors multiple ships	B1: Improper choice in emergency decision-making	C1: Lack of ship perception
A2: Insufficient vigilance	B2: Lack of experience in emergency disposal	C2: Situational awareness defect
A3 :Excessive fatigue	B3: Insufficient understanding of information	C3: Psychological difference

A4: Information overload	B4: No consideration to weather, sea conditions, etc.	C4: Uncoordinated man-machine interaction
A5: Insufficient sense of responsibility		C5: Insufficient training
A6: Poor physical and mental conditions		
A7: Automation-induced complacency		

In addition to the 16 nodes based on these human factors, it is necessary to use three additional nodes, namely "untimely perception", "incorrect decision" and "operation failure". These additional nodes indicate that the occurrence of a series of factors at each stage leads to the occurrence of relevant nodes at the same stage. Therefore, there are a total of 19 Bayesian nodes, which are described in Table 2.

Table 2. Description of the nodes in the proposed model

	Description		Description
A	Untimely perception	*B2*	Lack of experience in emergency disposal
A1	Negligence when one person monitors multiple ships	*B3*	Insufficient understanding of information
A2	Insufficient vigilance	*B4*	No consideration of weather, sea conditions, etc.
A3	Excessive fatigue	*C*	Operation failure
A4	Information overload	*C1*	Psychological difference
A5	Insufficient sense of responsibility	*C2*	Situational awareness defect
A6	Poor physical and mental conditions	*C3*	Lack of ship perception
A7	Automation-induced complacency	*C4*	Uncoordinated human-machine interaction
B	Decision failure	*C5*	Insufficient training
B1	Inappropriate emergency decision-making		

These nodes include the human error factors in the entire SCC, which means the "human" is not limited to only one operator but includes all the staff present in the SCC, i.e., monitoring personnel, helmsmen, cockpit operators, and so on.

The label *A1* refers to the negligence that occurs when one person is monitoring multiple ships. During the navigation of autonomous cargo ships, the responsibilities of the SCC staff are mainly concerned with monitoring the state of motion of the ships in real-time, which means monitoring multiple ships simultaneously during one session [27]. During the monitoring process, navigation information should be received continuously from each ship. Accordingly, when the volume of information handled by a staff member reaches a saturation value, known as “information overload” and labeled as *A4*, there is a possibility of negligence. In this context, "information overload" (*A4*) is the parent node of "negligence when one person monitors multiple ships" (*A1*).

Insufficient vigilance, labelled as *A2*, refers to the inability to perceive danger warning due to reduced vigilance by the staff present in the SCC towards monitoring of autonomous cargo ships. The “excessive fatigue”, labelled as *A3*, “insufficient sense of responsibility”, labelled as *A5*, and “poor physical and mental conditions”, labelled as *A6*, are all caused by “insufficient vigilance” (*A2*). In addition, the convenience arising due to automation also makes SCC personnel "over-dependent on automation", labelled as *A7*, thereby reducing personnel vigilance. Furthermore, when “insufficient vigilance” (*A2*) occurs among personnel, the above human factors are already included in the node, thus no separate statistics and illustrations will be given for the four nodes corresponding to *A3*, *A5*, *A6* and *A7*.

As for the inappropriate emergency decision-making, labelled as *B1*, when the monitoring personnel receive the danger warning from an autonomous cargo ship, the decision-makers often have “insufficient understanding of information", labelled as *B3*, during the process of emergency decision-making. The reason is due to different locations of the autonomous cargo ship and the personnel, or failure of the personnel to take into account the weather and sea conditions at the time of autonomous cargo ship navigation, which can lead to wrong decisions.

When it comes to “lack of experience in emergency disposal”, labelled as *B2*, the crew at the SCC need to acquire new skills for remotely managing the emergencies. This training can provide practical experience and help in avoiding incorrect decisions in response to remote emergencies.

The psychological difference is labelled as *C1*. An example of this difference is the inability of the operators in the SCC to acquire the real "ship perception", labelled as *C3*, since these operators operate on simulators. Thereby, real immersion in a scene cannot take place because of the simulated scenes, leading to "situational awareness defect", labelled as *C2* [28]. This situation results in a psychological gap for the operator who finds it unable to immerse himself in the scene, known as "uncoordinated man-machine interaction" (*C4*), which leads to operational failure.

In terms of "insufficient training" (*C5*), a group of new crews should not only master navigation technology, but also software equipment and algorithm-related knowledge. In other words, the requirements for crew quality are becoming stricter. Substandard operation technology is a major cause of shipwrecks. Therefore, the problem of insufficient training will be one of the most important reasons for operation failures in future navigation of autonomous cargo ships. To avoid these failures, the personnel should be required to undergo a gradually increasing amount of training.

3.3.3 Model structure

It can be observed that "insufficient vigilance" (*A2*) in the perception stage serves as the sub-node of four nodes, i.e., "excessive fatigue" (*A3*), "insufficient sense of responsibility" (*A5*), "poor physical and mental conditions" (*A6*), and "automation-induced complacency" (*A7*). Furthermore, it serves as the parent node of "negligence when one person monitors multiple ships" (*A1*) and "untimely perception" (*A*), while "automation-induced complacency" is also the parent node of "negligence when one person monitors multiple ships" (*A1*).

In the decision stage, "inappropriate decision" (*B1*) serves as the sub-node of "insufficient understanding of information" (*B3*) and "no consideration to weather, sea conditions, etc." (*B4*), while both *B1* and "lack of experience in emergency disposal" (*B2*) are the parent nodes of "decision failure" (*B*).

In the operation stage, "psychological difference" (*C1*) serves as the sub-node of "situational awareness defect" (*C2*) and "lack of ship perception" (*C3*). Both *C1* and *C2* are the parent nodes of "uncoordinated man-machine interaction" (*C4*). Meanwhile, *C4* and "insufficient training" (*C5*) are the parent nodes of "operation failure" (*C*). Based on these relationships between children and parent nodes, the three-stage BNs model can be constructed, as shown in Figs. 4-6.

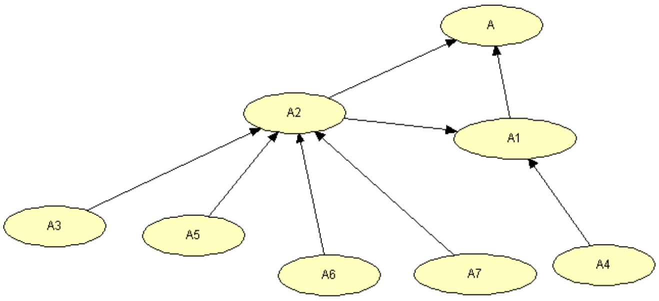


Figure 4. Bayesian Network model of the perception stage (Notes: *A* - Untimely perception; *A1* - Negligence when one person monitors multiple ships; *A2* - Insufficient vigilance; *A3* - Excessive fatigue; *A4* - Information overload; *A5* - Insufficient sense of responsibility; *A6* - Poor physical and mental conditions; *A7* - Automation-induced complacency)

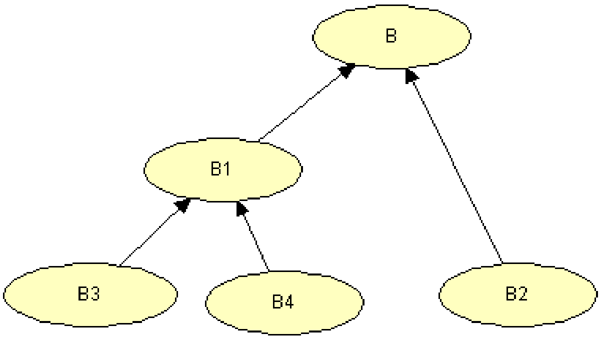


Figure 5 Bayesian Network model of the decision stage (NOTES: B - Decision failure; B1 - Inappropriate emergency decision-making; B2 - Lack of experience in emergency disposal; B3 - Insufficient understanding of information; B4 - No consideration to weather, sea conditions, etc.)

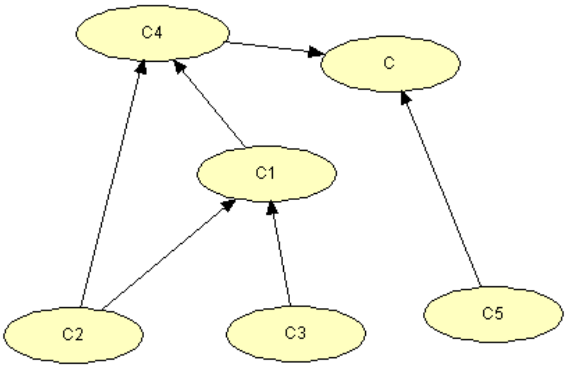


Figure 6. Bayesian Network model of the operation stage (Notes: C - Operation failure; C1 - Psychological difference; C2 - Situational awareness defect; C3 - Lack of ship perception; C4 - Uncoordinated human-machine interaction; C5 - Insufficient training.)

4 CASE STUDY

4.1 Data description

As the availability of data related to human factors in the SCCs is limited, expert experience method is adopted for analysis of basic occurrence probability of these factors. On the other hand, the human factors in this study provide theoretical support, risk prevention and control measures for future construction of the SCCs and personnel training. In this section, the expert data are processed by fuzzy triangular numbers. For instance, three well-known experts in the field of water safety evaluation and having experience for more than 15 years, were invited to provide evaluation comments on basic event probability of human error in autonomous cargo ship navigation. Considering the rich working experience of the experts, their comments were deemed as important as other methods described in the following steps:

- Frequency grading

  In the process of risk assessment, it is sufficient to only use frequency for event grading[29], such as the grading method provided in Table 3. This frequency can be either a risk assessment indicator or a safety performance indicator. It can be observed based on the standard and definition of the frequency level that the frequency in a level is usually 10 times higher than that in the previous level. According to this definition, the corresponding level number can also be approximated on a logarithmic scale.

Table 3. Accident frequency level

Level	Frequency / Year	Description
Very low	10^{-5}~0	Events that are almost impossible
Low	10^{-5}~10^{-3}	Very rare events, not seen in similar projects.
Medium	10^{-3}~10^{-1}	Rare events, people may encounter in their lifetime
High	0.1~1	An event that has happened, whose reoccurrence in the future is normal
Very high	1~10	Events expected to happen frequently

- Processing of fuzzy probability

  The membership function of the trigonometric function is as follows:

$$f(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{m-a} & a \le x \le m \\ \dfrac{b-x}{b-m} & m \le x \le b \\ 0 & x > b \end{cases}. \tag{5}$$

It can be seen from (5) that the triangular number can be represented by three parameters, i.e., *a*, *m* and *b*. In order to generate the experts' score with reference to Table 3, five semantic values shown in Table 4 are specified to represent different fuzzy numbers. The membership function of the corresponding triangular fuzzy number is shown in Fig. 7.

Table 4. Semantic values of event occurrence probability and the corresponding triangular fuzzy number

No.	Semantic value	Triangular fuzzy number
1	Very low	(0,0,0.3)
2	Low	(0,0.3,0.5)
3	Medium	(0.3,0.5,0.7)
4	High	(0.5,0.7,1)
5	Very high	(0.7,0.7,1)

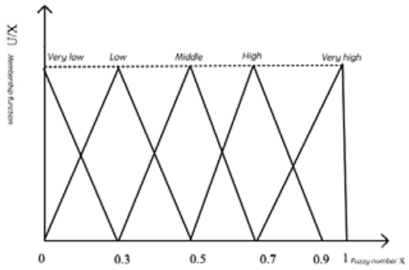


Figure 7. Membership function of fuzzy triangular numbers

- Analysis sequences

It is necessary to synthesize the semantic judgments of multiple experts for a more accurate characterization of the event occurrence possibility through fuzzy numbers. This paper adopts the fuzzy number synthesis method using a weighted summation, and using this method, the comprehensive evaluation of an event *i* can be expressed as

$$M_i = W_1 F_{1i} + W_2 F_{2i} + \cdots + W_n. \tag{6}$$

The weight value of the j^th^ expert (j = 1, 2,…,n) can be represented by W_j, and F_{ji} represents the semantic evaluation fuzzy number of the j^th^ expert for the i^th^ event, (i = 1,2,…,m).

(1) Deburring

In this paper, the mean area method is used to process the fuzzy probability and obtain the exact probability [30]. The formula is shown as follows:

$$p'_n = \frac{a'_n + 2m'_n + b'_n}{4}. \tag{7}$$

(2) Probability normalization

For each basic event, the sum of the state probabilities must be equal to one, therefore, the probability given by (4) should be normalized as follows: [31-32]

$$p_n = \frac{p'_n}{4\sum_{j=o}^{n-1} p'_n}. \tag{8}$$

The final probability of the root node obtained after deburring is shown in Table 5.

Table 5. Basic probability of the root node

Variable	A3	A5	A6	A7	A4	B2
Probability	1.62e-3	1.64e-3	1.66e-3	1.86e-3	2.64e-3	1.62e-3
Variable	B3	B4	C2	C3	C5	
Probability	1.61e-3	1.75e-3	2.23e-3	2.45e-3	1.62e-3	

4.2 CPT estimation

The conditional probability table for each sub-node can be determined based on the root node probability. The SCC can face multiple uncertainties during its construction due to lack of available data. This section uses both expert interviews and questionnaires to obtain the conditional probability table for the nodes, where the interview questions are mainly based on probability assignment. Based on the given constraints, the interviewed experts will independently give the corresponding probability values, which are then statistically analyzed to obtain an average value.

Taking the sub-node "insufficient vigilance" (*A2*) as an example, as shown in Figs. 4 and 5, there are four parent nodes of A2, including "excessive fatigue" (*A3*), "insufficient sense of responsibility" (*A5*), "poor physical and mental conditions" (*A6*) and "over-reliance on automation" (*A7*), where *A2* can have a value of either zero or one. The former value indicates that the operator is not vigilant enough, while the latter value indicates that the operator is vigilant. Similarly, the four parent nodes of *A2* also have two states, namely zero and one. The conditional probability table of the sub-node "insufficient vigilance" (*A2*) with respect to other states is shown in Table 6.

Table 6. Conditional Probability of "Insufficient Vigilance" (*A2*)

A3	A5	A6	A7	A2	
				Y	N
Y	Y	Y	Y	0.0212	0.9788
			N	0.0141	0.9859
		N	Y	0.0196	0.9804
			N	0.0136	0.9864
	N	Y	Y	0.02	0.98
			N	0.0135	0.9865
		N	Y	0.0198	0.9802
			N	0.0137	0.9863
N	Y	Y	Y	0.0178	0.9822
			N	0.0111	0.9889
		N	Y	0.019	0.981
			N	0.0128	0.9872
	N	Y	Y	0.0188	0.9812
			N	0.0123	0.9877
		N	Y	0.01	0.99
			N	0.0038	0.9962

4.3 Results

Based on the three-stage BNs topology structure of human factors in the SCC, the knowledge of experts is effectively extracted using the calculation method described in the previous sub-section. Subsequently, the conditional probability table of each node is calculated and the probability values are input to the analysis software. The network prediction model for the three stages are shown in Figs. 8-10 and the human error occurrence probabilities *P1*, *P2*, *P3* in each stage can be obtained using these models. Using these probability values in (4), the occurrence probability of ship accident, also called the total human error probability in emergency situations, is calculated as $P=1-0.9961\times0.9984\times0.9969=8.58\text{e-}3$.

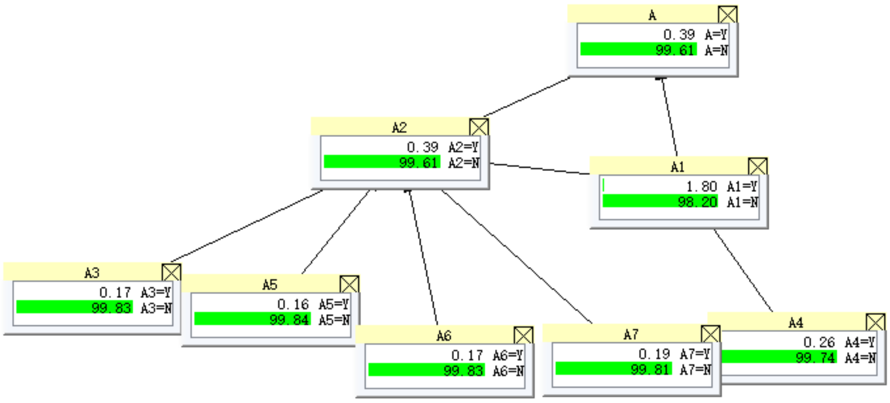


Figure 8. Topology network in the perception stage

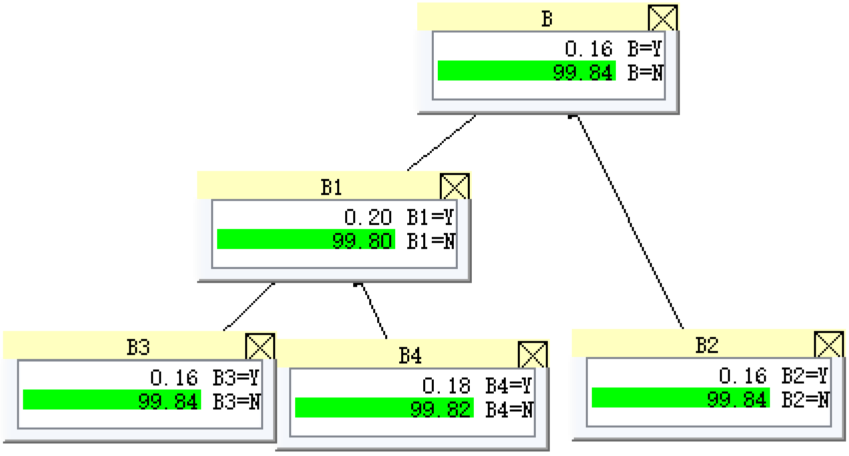


Figure 9. Topology network in the decision stage

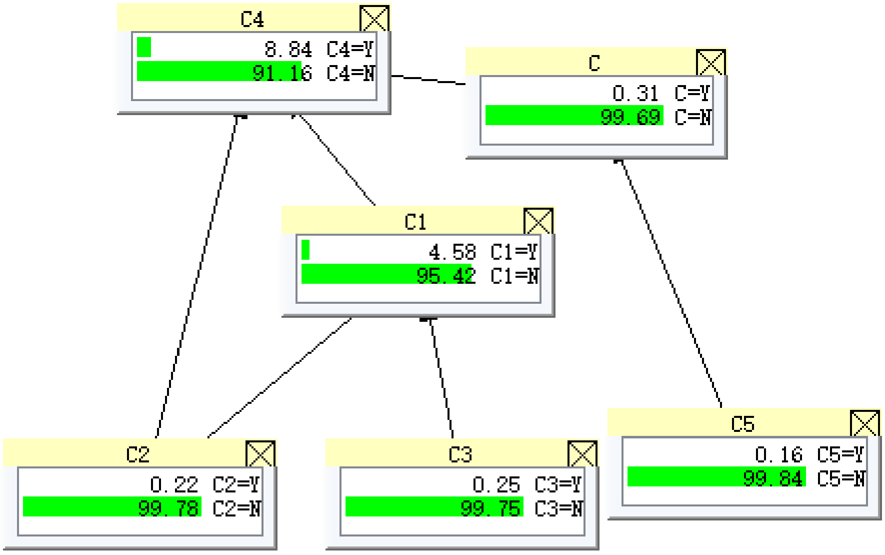


Figure 10. Topology network in the operation stage

4.4 Sensitivity analysis

In order to analyse the influence of each human factor on autonomous cargo ship navigation accidents, the sensitivity of the proposed BNs model of the SCC is analysed in this section. First, the occurrence probability of each parent node is assigned the value of one, i.e., $P_{(Cij=1)} = 100\%$, where i denotes the node category of the risk factor and j denotes the node number of the risk factor. Then, a full probability variation table of risk events in autonomous cargo ships caused by human errors in the SCC is obtained after prediction.

Taking the example of node “excessive fatigue” (*A3*), the network topology diagram of the perception stage variations when the monitoring staff is excessively fatigued is shown in Fig. 11.

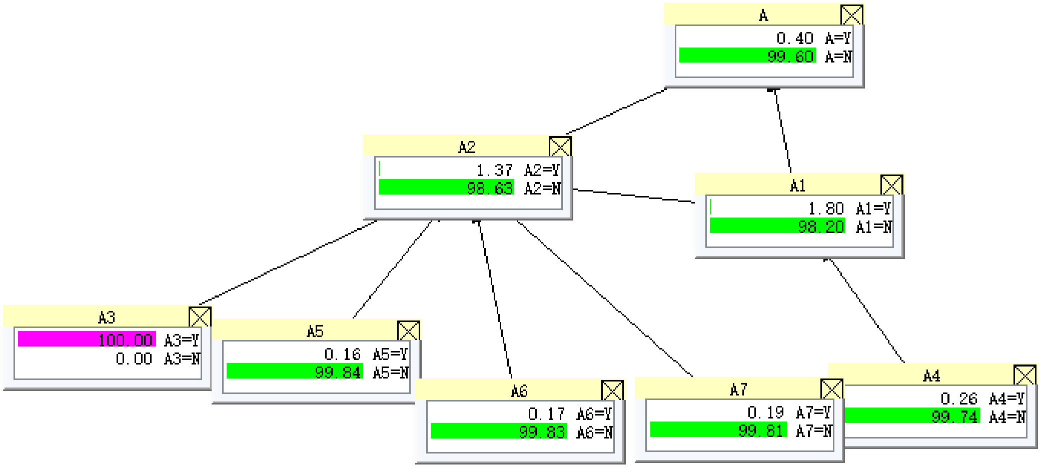


Figure 11. Schematic diagram of the calculation results of the posterior probability in the perception stage

Based on (4) and Fig. 11, the total occurrence probability of accidents can be calculated as P=8.68e-3 for the case of "excessive fatigue" (*A3*) in emergency scenario. Similarly, the full occurrence probability of accidents relative to each node variable can be predicted, as shown in Table 6.

Table 7. Sensitivity analysis of human error factors in the SCC

Variable	Posterior occurrence probability
A1	2.46e-2
A2	1.88 e-2
A3	8.68e-3
A4	1.99 e-2
A5	8.67e-3
A6	8.63e-3
A7	9.73e-3
B1	1.49 e-2
B2	1.93 e-2
B3	9.02 e-3
B4	8.98 e-3
C1	9.48 e-3
C2	2.05 e-2
C3	1.80 e-2
C4	2.29 e-2
C5	1.81 e-2

Table 6 shows the posterior probability of each node. The sensitivity of human factors affecting the autonomous cargo ship navigation accidents is ranked as follows:

A1> C4> C2> A4> B2> A2> C5> C3> B1 > A7> C1> B3 > B4> A3> A5> A6.

4.5 Model validation

This paper could be observed that when the staff on the SCC has to deal with the emergency disposal of autonomous cargo ships in section 4.4. The factors whose posterior probabilities are higher than the prior probability include “negligence when one person monitors multiple ships”, “uncoordinated man-machine interaction”, “situational awareness defect”, “information overload”, “lack of experience in emergency disposal”, “insufficient vigilance” and “insufficient training”, with a combined probability value of greater than 100%. In fact, in the whole system, "negligence when one person monitors multiple ships" and "uncoordinated man-machine interaction" have the two highest node sensitivities, which significantly impact the occurrence of ship accidents. In other words, these two human error factors are highly likely to cause ship accidents due to the failure in personnel emergency disposal.

Although existing studies focused mainly on human factor identification for autonomous cargo ships, they lacked details about different human error types and their importance in the emergency response disposal by the SCC. For example, Ramos explored human factors in the navigation process of autonomous cargo ships. This study mainly used the event tree analysis to analyse which human error may occur in the ship control and its degree of impact on consequent accidents, based on the progressive order of events. In addition, another study [29] figured out that the most important human errors affecting ship navigation include personnel negligence, information overload, situational awareness defect, skill degradation and insufficient vigilance caused by ignorance. The study emphasized the human factors such as monitoring personnel’s negligence and situational awareness defect, which is consistent with the human error factor ranking presented in this study.

5 DISCUSSION

This study utilized THERP and Bayesian theory to predict human error probabilities in emergency disposal when a ship is controlled remotely by the SCC. The findings manifested that the probability of error by the operator in the SCC during the emergency process was 8.58e-3, which is slightly higher than that of traditional ships. It was observed by a study of existing literature that the researchers are not optimistic about the safety of autonomous cargo ships, because although the human safety is guaranteed when the operators are transferred from the ship to the SCC, the risk index for the ship itself is higher than that of traditional ships. Therefore, there is an urgent need for further research on the safety of autonomous cargo ships.

Based on the human factors sensitivity results in case of emergency disposal of autonomous cargo ships, an analysis of eight risk factors having a high sensitivity score was carried out. The analysis revealed that it was necessary to strictly control the “negligence when one person monitors multiple ships” (*A1*). Similarly, the problem of "information overload" (*A4*) should also be avoided. To manage "uncoordinated human-system interaction" (*C4*), "situational awareness defect" (*C2*) and "lack of ship perception" (*C3*), it is necessary to have realistic simulations and training, while an emergency plan system should be improved to deal with “lack of experience in emergency disposal” (*B2*). Finally, crew training should be strengthened to avoid "insufficient vigilance" (*A2*) and “insufficient training” (*C5*).

In summary, there are several points that the clients should pay attention to when constructing the SCCs and training the operators. These points include: “standardize the number of ships monitored by one person”, continuously “enhance truthfulness of simulated cabins”, strengthening “emergency plan improvement and emergency disposal drills” and mitigating "insufficiency of education and training". These points can provide theoretical basis and reference opinions, thereby reducing human errors in emergency disposal of autonomous cargo ships.

6 CONCLUSION AND FUTURE WORK

This paper analysed the emergency disposal process in the SCC of autonomous cargo ship with the third degree of autonomy. Based on this analysis, the human error probability was divided into three stages of perception-decision-execution, using the human error probability prediction method. Then, the BNs models of the three stages were constructed, followed by calculations of the basic probability of root node based on processing expert opinions using triangular fuzzy numbers. Subsequently, the conditional probability of each intermediate node in the network was also determined, which was used to obtain the human error probability of the entire emergency treatment process. This probability was equal to 8.58e-3. Finally, the factor importance was ranked based on sensitivity analysis of each human factor. Specifically, the top eight risk factors included “negligence when one person monitors multiple ships” (*A1*), “uncoordinated human-system interaction” (*C4*), “situational awareness defect” (*C2*), “information overload” (*A4*), “lack of experience in emergency disposal” (*B2*), "insufficient vigilance" (*A2*), "insufficient training" (*C5*) and "lack of ship perception" (*C3*). Additionally, Risk Control Options were proposed to provide theoretical suggestions and support for the future construction of SCCs, and staff training.

As the concept of SCC is still in the design stage, the human factor error model for the SCC needs further improvement. The predicted human error probability and conclusions in this paper only serve as a reference for designing a SCC, and risk prevention and control. The human error model can be further studied in future when the SCCs become operational.

ACKNOWLEDGMENTS

The research was supported by the National Key Research and Development Project (2017YFC0804900, 2017YFC0804904), National Key Technologies Research & Development Program (2017YFE0118000). This work has also been supported by the H2020-MCSA-RISE project “RESET - Reliability and Safety Engineering and Technology for large maritime engineering systems” (No.730888).

REFERENCES

[1] International Maritime Organization (IMO). [2018-10-19]. http://www.imo.org/en/MediaCentre/PressBriefings/Pages/08-MSC-99-MASS-scoping.aspx

[2] Maritime Safety Committee (MSC). [2018-10-19]. http://www.imo.org/en/MediaCentre/MeetingSummaries/MSC/Pages/MSC-99th-session.aspx

[3] Burmeister C H, Bruhn W C, Rødseth, Ø J, Porathe, T. Can unmanned ships improve navigational safety?. Transport Research Arena 2014, Paris, 2014.

[4] H-C Burmeister, W Bruhn, Ø J Rødseth, T Porathe. Autonomous Unmanned Merchant Vessel and its Contribution towards the e-Navigation Implementation: The MUNIN Perspective. International Journal of e-Navigation and Maritime Economy, 2014, 1: 1-13.

[5] Ren J, Jenkinson I, Wang J, et al. A methodology to model causal relationships on offshore safety assessment focusing on human and organizational factors. Journal of safety research, 2008, 39(1):87-100.

[6] Porathe T, Prison J, Man Y. Situation awareness in remote control centres for unmanned ships. Human Factors in Ship Design & Operation, London: UK, 2014.

[7] Ahvenjärvi S. The human element and autonomous ships. International Journal on Marine Navigation and Safety of Sea Transportation, 2016, 10 (3): 517-521.

[8] Ramos M A, I., et al. Accounting for human failure in autonomous ship operations. European Safety and Reliability Conference, 2018.

[9] Trudi H, Samrat G. Autonomous merchant vessels: examination of factors that impact the effective implementation of unmanned ships. Australian Journal of Maritime And Ocean Affairs, 2016, 8(3):206-222.

[10] Porathe T, Prison J, Man Y. Situation awareness in remote control centres for unmanned ships. Human Factors in Ship Design & Operation, London: UK, 2014.

[11] K. Wróbel, J. Montewka, P. Kujala. Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. Reliab Eng Syst Saf, 165 (September) (2017), pp. 155-169

[12] Y. Man, M. Lundh, T. Porathe, S. MacKinnon. From desk to field - Human factor issues in remote monitoring and controlling of autonomous unmanned vessels. Procedia Manufacturing, 6th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences (AHFE), 3 (2015). January 26, 74–81

[13] M. Wahlström, J. Hakulinen, H. Karvonen, I. Lindborg. Human factors challenges in unmanned ship operations – Insights from other domains. Procedia Manufacturing, 6th International Conference on Applied Human Factors

and Ergonomics and the Affiliated Conferences (AHFE)., 3 (2015), pp. 1038-1045 January

[14] T. Hogg, S Ghosh. Autonomous merchant Vessels: examination of factors that impact the effective implementation of unmanned ships. Austr J Marit Ocean Affairs, 8 (3) (2016), pp. 206-222

[15] Valdez Banda, O.A., Goerlandt, F., Kujala, P., Montewka, J. Expert elicitation of Risk Control Options to reduce human error in winter navigation. Safety and Reliability of Complex Engineered Systems - Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015, Pages 3137-3146

[16] Ung, S-T. Evaluation of human error contribution to oil tanker collision using fault tree analysis and modified fuzzy Bayesian Network based CREAM. Ocean Engineering, Volume 179, 1 May 2019, Pages 159-172.

[17] Macrae, C. Human factors at sea: Common patterns of error in groundings and collisions (2009) Maritime Policy and Management, 36 (1), pp. 21-38.

[18] Nevalainen, M., Helle, I., Vanhatalo, J. Estimating the accute impacts of Arctic marine oil spilss using expert elicitation. Marine Pollution Bulletin. Vol. 131. June 2018, Pages 782-792

[19] Ren, J., et al. (2008). A methodology to model causal relationships on offshore safety assessment focusing on human and organizational factors. Journal of Safety Research, 39 (1), 87–100.

[20] Ren, J.,et al., (2009). An offshore risk analysis method using fuzzy Bayesian Network. Journal of Offshore Mechanics and Arctic Engineering, 131 (4), 3–16.

[21] Stanton N. Human Factors in Nuclear Safety. London(UK): Taylor & Francis Ltd, 1996.

[22] Marilia A. Ramos, Ingrid B. Utne, Ali Mosleh. On factors affecting autonomous ships operators performance in a SCC. Probabilistic Safety Assessment and Management PSAM 14. Los Angeles: CA, 2018.

[23] Zhang J, Zhang D, Yan X P, et al. A distributed anti-collision decision support formulation in multi-ship encounter situations under COLREGs. Ocean Engineering, 2015(105):336-348.

[24] Zhou, T., Wu, C., Zhang, J., and Zhang, D. Incorporating CREAM and MCS into Fault Tree analysis of LNG carrier spill accidents. Safety Science, 2017(96):183-191.

[25] Zhang MY, Zhang D, Goerlandt F, Yan XP, Kujala P. Use of HFACS and fault tree model for collision risk factors analysis of icebreaker assistance in ice-covered waters. Safety Science. 2019;111:128-43.

[26] Ahvenjärvi S. The human element and autonomous ships. International Journal on Marine Navigation and Safety of Sea Transportation, 2016, 10 (3): 517-521.

[27] Marilia A. Ramos, Ingrid B. Utne, Ali Mosleh. On factors affecting autonomous ships operators performance in a SCC. Probabilistic Safety Assessment and Management PSAM 14. Los Angeles: CA, 2018

[28] Man Y M, Lundh M, Porathe T. Seeking harmony in shore-based unmanned ship handling - from the perspective of human factors, what is the difference we need to focus on from being onboard to onshore?. Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE, Kraków, 2014.

[29] ABS Consulting. Marine safety: tools for risk-based decision making. Rockville, MD: Rowman & Littlefield, 2002.

[30] Swain A D and Guttmann H E. Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Application. NUREG/CR-1278, 1983.

[31] Abujaafar K M, Qu Z, Yang Z, et al. Use of evidential reasoning for eliciting Bayesian subjective probabilities in human reliability analysis. 2016 11th System of Systems Engineering Conference (SoSE). IEEE, 2016: 1-6.

[32] Zhang D, Yan X P, Yang Z L, et al. Incorporation of formal safety assessment and Bayesian network in navigational risk estimation of the Yangtze River. Reliability Engineering & System Safety, 2013, 118:93-105.

Comparison of system modelling techniques for autonomous ship systems

Sunil Basnet*, Osiris A. Valdez Banda, Meriam Chaal, Spyros Hirdaris and Pentti Kujala
Aalto University, School of Engineering, Marine Technology Group, Espoo, Finland

ABSTRACT

As autonomous ships are currently developed, modern technologies are implemented into ship systems for enabling autonomous operations. Tight coupling in safety-critical systems created new challenges for the engineers and operators. Designing, operating and analyzing these complex systems requires a deep understanding about the system composition, requirements and expected behavior or functionality. The increasing complexity of the systems requires the implementation of modern model-based approaches. Instead of large texts, these new modelling techniques aim to present detailed system information with simplified models. This paper compares system modelling techniques known as System Modelling Language (SysML) and Object Process Methodology (OPM). These methods are used to model a Dynamic Positioning system (DP-system). Results show that the SysML is more suitable than OPM for modelling the autonomous ship systems due to its ability to present detailed system information in a simple and coherent way.

Keywords: Modeling methods; System Modeling Language; Object Process Methodology; Autonomous ship system; Dynamic Positioning System

1 INTRODUCTION

Numerous maritime industry stakeholders are currently exploring the options for developing new autonomous ship concepts (MUNIN, 2016). To date, research and engineering efforts focus on the development of technologies that could enable safe autonomous ship operations. However, the increased functionalities of autonomous engineering systems onboard ships are supported by advanced software and enable complex operations (Levander, 2017). This increased complexity represents new challenges for engineers in the system's development, operation and management. These challenges link with three keystreams of innovation: (a) the management of system complexity; (b) the understandability of the system, and (c) the communication of the gathered information (Holt & Perry, 2019). The solutions to handle these issues need to have a consistent interconnection. This need for interconnection is evident as the absence of complexity management generates a difficult process for gathering information and results in an inefficient communication of the system information. This is the reason why there is a need for new alternatives for the systemic handling of complexity in modern autonomous systems and associated operations. These alternatives should support engineers and operators in the processing of system information. Moreover, the alternatives must effectively communicate and utilize the system information throughout diverse engineering processes such as development, analysis, operation and maintenance (Holt & Perry, 2019).

Advancement of technologies and their complexity requires a modern model-based approach (Grobshtein et al., 2007). Accordingly, engineers and operators must understand how system components interact with each other and surrounding systems (Weck et al., 2011). Novel methods based on Model Based Systems Engineering (MBSE) and System of Systems Engineering (SOSE) could offer solutions for the modelling of the complex modern systems. The International Council on Systems Engineering (INCOSE) defines MBSE as "the formalized application of modeling to support system requirements, design, analysis, verification and validation. It begins at the conceptual design phase and continues throughout development and later life cycle phases" (Friedenthal et al., 2007). Most of the methods based on MBSE are supported with computer tools for handling the system complexity. These enable a systematic organization of the modeling

---

[*] Sunil Basnet: +358451812811, Sunil.basnet@aalto.fi

process. MBSE implements the principles of SOSE for a holistic modeling that begins with a general system level and proceed to subsystems and components. This holistic modeling includes system descriptions that are effectively communicated by diagrams and text.

The models providing the system information have been utilized for various purpose in systems safety engineering. Some analysis methods such as the System's Theoretic Process Analysis (STPA) and the Functional Resonance Analysis Method (FRAM) utilize models to guide the analysis process. For example, in STPA, the so-called safety control structure model provides system information about the controllers, controlled processes and the interactions between them. This information is then used to define the scenarios where the system can be in an unsafe state (Leveson & Thomas, 2018). On the other hand, in FRAM, the model presents functional system interactions using different aspects such as requirements, resources, control, time, input and output. This system information is then utilized to understand the system's couplings and performance variability (Hollnagel, 2012). Thus, the models providing the system information have a crucial role when conducting a system analysis. With these purposes, models have been effectively utilized in several domains for enhancing system design decision making (Russell, 2012), system development (D'Ambrosio & Soremekun, 2017), Safety analysis (Mhenni et al., 2013) and Security analysis (Best et al., 2007).

This paper compares two state of the art modeling methods based on MBSE and SOSE principles knowns as System Modelling Language (SysML) and Object Process Methodology (OPM). The paper includes a case study where a Dynamic Positioning (DP) system is modeled to compare the funcitonality of these methods. The DP-system is selected for this study because it is considered to be one of the main systems where autonomy in maritime operations is critical in terms of both positioning and navigation. Based on the generated models, the similarities and differences of these methods are compared. Finally, the applicability of these methods in autonomous ship systems is discussed.

2 METHODS

2.1 System Modelling Language (SysML)

SysML is a modelling language developed by Object Management Group (OMG) in 2007. It was derived from UML (Unified Modeling Language), which is widely used in Software Engineering (Friedenthal et al., 2015). SysML has been recognized by the International Standards Organization (ISO) since 2017 (ISO/IEC 19514:2017). OMG defines SysML as "a general-purpose graphical modelling language for specifying, analyzing, designing, and verifying complex systems that may include hardware, software, information, personnel, procedures, and facilities. It is an enabler of a MBSE approach to improve productivity, quality, and reduce risk for complex systems development." (Object Management Group, 2017). SysML contains 9 different types of diagrams with each of them having a specific purpose, different level of abstraction and providing a different view of the same system. These diagrams aim to model the structural and behavioral aspects of the system and are classified as: (Holt & Perry, 2019):

I. **Structural diagrams:**

   a. The **block definition diagram**, labelled *bd,* presents the hierarchy of system elements (systems, sub-systems and components) to define their properties. For each element, a block is created with a possibility to add any related properties or specifications as appropriate such as weight, dimensions and cost.
   b. The **internal block diagram**, labelled *ibd,* presents the internal structure of the system or sub systems. It presents how the parts are interconnected from an inside perspective by adding parts, ports and connectors.
   c. The **requirements diagram**, labelled *req,* is used to list the requirements of the system elements, which can include guidelines, standards, rules etc. During the system development life cycle, this diagram allows engineers to verify and validate these requirements.
   d. The **parametric diagram**, labelled *par,* is used for conducting engineering analysis as well as verification and validation of the system requirements. The

constraints for the system analysis are defined and then the performance of systems is calculated and evaluated using the property values defined in system elements block. It can include various engineering analysis such as trade studies, sensitivity analysis, performance analysis and design optimization (Friedenthal et al., 2015).

e. The **package diagram**, labelled *pkg,* is used to organize all SysML diagrams. The diagrams with similarities are identified and grouped in a package (folder) which can be referenced with a unique name as suitable by the modeler. Furthermore, if a diagram belongs to two different groups then an import relationship can be applied to import the diagram or diagram elements from one package to another.

II. **Behavioral diagrams:**

a. The **activity diagram**, labelled *act,* presents the flow-based behavior of the system when performing certain activities. It shows the flow of control, flow of objects, input required for executing the activity and output that the activity produces.
b. The **sequence diagram**, labelled *sd,* shows the sequential flow and exchange of messages between different system elements when they interact with each other.
c. The **use case diagram**, labelled *uc*, describes the functionality of the system by presenting the actors (users of the system) and tasks required to execute the function. Furthermore, the operational requirements of the system can be refined using the use case diagram, which shows how the system functions fulfil the system requirements.
d. The **state machine diagram**, labelled *stm,* presents the behavior of the system elements when the state transitioning occurs. It is used to describe the state-dependent behavior of the system elements during the system operation.

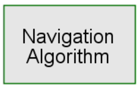
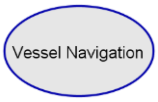
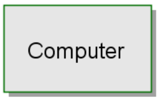
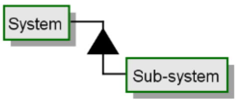
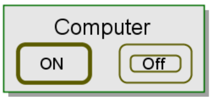
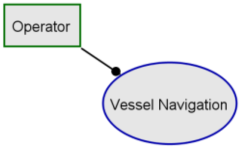
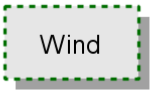
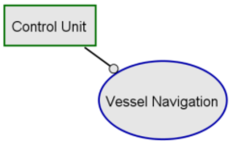
2.2 Object Process Methodology (OPM)

OPM is a system modelling paradigm introduced by Prof. Dov Dori in 1995 (Dori, 1995). It is a holistic approach, which models a system's structural and behavioral aspects in a single and unified diagram. It can support engineers during system design, development, maintenance, and effective communication (Weck et al., 2011). In 2015, ISO recognized OPM as an international standard modeling language for producing conceptual models of a system (ISO/PAS 19450:2015).

OPM focuses on three entities that are inherent in a system namely objects, processes and the links in between them. Dori (2002) defines objects as the "things that exist in the system" and consequently, he defines processes as the "things that happen in the system. For presenting these system descriptions, it uses both the graphical form via Object-Process Diagrams (OPDs) and textual form via Object-Process Language (OPL).

The OPD's consist of one System Diagram (SD) and many In-Zoomed diagrams. The SD presents the system level (top level) elements, while the In-Zoomed diagrams present each of the elements of SD in higher detail. In this way engineers may start designing at abstract level and add refinements as needed. Each OPD includes a collection of sentences, in OPL format, (textual modality). OPL is defined by Dori (2016) as "a subset of English that expresses textually the OPM model that the OPD set expresses graphically". This feature creates human readable auto-generated texts that can be useful for engineers that prefer texts over graphics and can also be used to create technical specifications (Dori, 2016). Table 1 presents some symbols and their representation in OPDs.

Table 1: Some symbols and their representation in OPD's

Symbol	Representation	Symbol	Representation
	**Non-physical Object**: Flat rectangle		**Process**: Eclipse
	**Physical Object**: Shaded rectangle		**Aggregation -participation link**: Black filled triangle **OPL**: System consists of Sub-system
	**States**: rounded rectangle **Initial state (on):** thick border **Final state (off):** double border		**Agent link**: line with black filled circle **OPL:** Operator handles Vessel Navigation
	**Environmental Object**: with dashed border		**Instrument link**: line with white filled circle. **OPL**: Vessel Navigation requires Control Unit.

3 MODELING A SHIP DYNAMIC POSITIONING SYSTEM (DP-SYSTEM)

3.1 Dynamic Positioning system

The International Maritime Organization (IMO) defines Dynamically Positioned vessel (DP-vessel) as "a unit or a vessel which automatically maintains its position (fixed location or predetermined track) exclusively by means of thruster force". It also defines the Dynamic Positioning system (DP-system) as "the complete installation necessary for dynamically positioning a vessel comprising of a power system, thruster system and DP-control system" (IMO-MSC/Circular.645).

The DP-system estimates the required thrust and rudder angle to maintain the vessel position against wind, waves and current; and controls the engine, thrusters, propellers and rudders accordingly for reaching and maintaining the desired position. It consists of several position-reference units and sensors to estimate the vessel motion and position in addition to the forces affecting them. The sample DP-system, K-Pos DP-21 (Kongsberg, 2014), is modelled in the upcoming sections with a focus on automatic vessel positioning mode. This system satisfies the requirements of IMO equipment Class 2 as specified in IMO-MSC/Circular.645. The details of this system and its functions are available in Kongsberg (2014) and is used for modeling in Sections 3.2 and 3.3.

3.2 Modeling with SysML using Visual Paradigm tool

Figure 1 presents the package diagram for the DP-system. The package diagram consists of 4 main packages: requirements, behavior, structure and parametric, which are further refined into sub packages as shown in the figure.

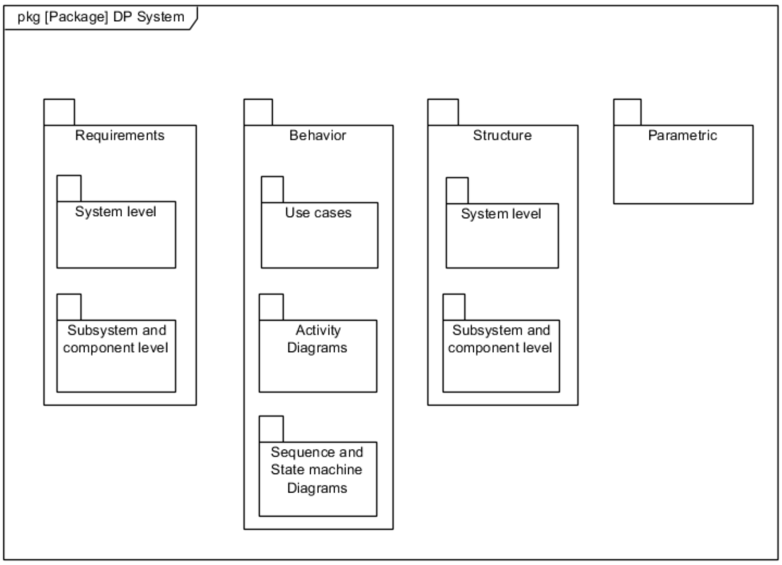


Figure 1: The package diagram for DP-system

Next, inside the requirements package, the requirements diagrams are created for the DP system, sub-systems and their components. Figure 2 presents the system level (DP-system) requirements from IMO MSC/Circular 645 for Class 2 vessels. Each of the requirements inside the diagram has a title (heading), text (describing the requirement) and ID (identity number). Furthermore, as shown in the requirement titled as “International Standard Units” in Figure 2, it can also include various fields such as verify method, risk involved due to unfulfilled requirements, and current status of the requirement verification.

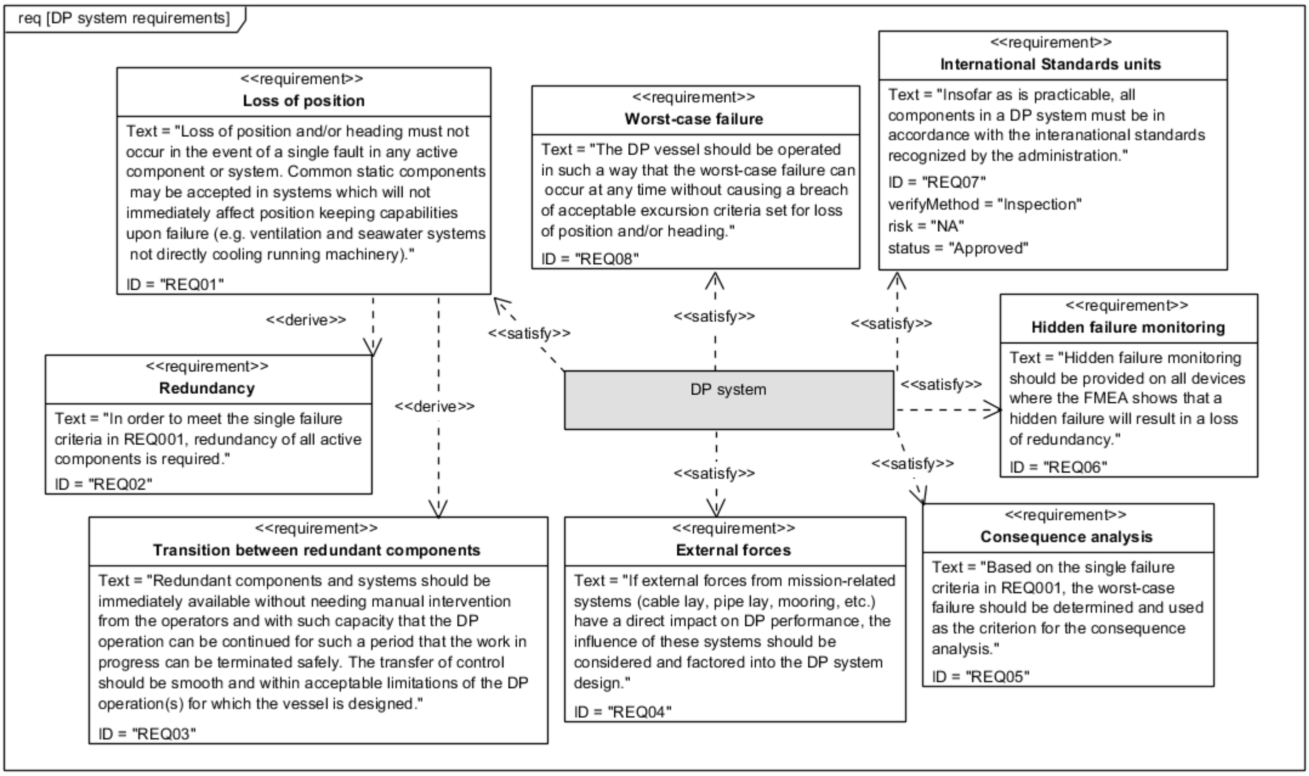

req [DP system requirements]
<<requirement>>
Loss of position
Text = "Loss of position and/or heading must not occur in the event of a single fault in any active component or system. Common static components may be accepted in systems which will not immediately affect position keeping capabilities upon failure (e.g. ventilation and seawater systems not directly cooling running machinery)."
ID = "REQ01"
<<requirement>>
Worst-case failure
Text = "The DP vessel should be operated in such a way that the worst-case failure can occur at any time without causing a breach of acceptable excursion criteria set for loss of position and/or heading."
ID = "REQ08"
<<requirement>>
International Standards units
Text = "Insofar as is practicable, all components in a DP system must be in accordance with the interanational standards recognized by the administration."
ID = "REQ07"
verifyMethod = "Inspection"
risk = "NA"
status = "Approved"
<<derive>>
<<satisfy>>
<<satisfy>>
<<satisfy>>
<<requirement>>
Hidden failure monitoring
Text = "Hidden failure monitoring should be provided on all devices where the FMEA shows that a hidden failure will result in a loss of redundancy."
ID = "REQ06"
<<requirement>>
Redundancy
Text = "In order to meet the single failure criteria in REQ001, redundancy of all active components is required."
ID = "REQ02"
<<derive>>
DP system
<<satisfy>>
<<satisfy>>
<<satisfy>>
<<requirement>>
Transition between redundant components
Text = "Redundant components and systems should be immediately available without needing manual intervention from the operators and with such capacity that the DP operation can be continued for such a period that the work in progress can be terminated safely. The transfer of control should be smooth and within acceptable limitations of the DP operation(s) for which the vessel is designed."
ID = "REQ03"
<<requirement>>
External forces
Text = "If external forces from mission-related systems (cable lay, pipe lay, mooring, etc.) have a direct impact on DP performance, the influence of these systems should be considered and factored into the DP system design."
ID = "REQ04"
<<requirement>>
Consequence analysis
Text = "Based on the single failure criteria in REQ001, the worst-case failure should be determined and used as the criterion for the consequence analysis."
ID = "REQ05"

Figure 2. The requirement diagram for DP system

Then, the diagrams presenting the structural aspects of the DP-system are created. Figures 3 and 4 present the block definition diagram of the DP-system domain and DP control unit respectively. Each of the system elements are modeled as a block and are connected with a composite association (whole-part relationship). Furthermore, each of the components block can include relevant properties and values. For example, the “operator station” block in Figure 6 includes fields where the information about its current, voltage, dimension and weight can be placed.

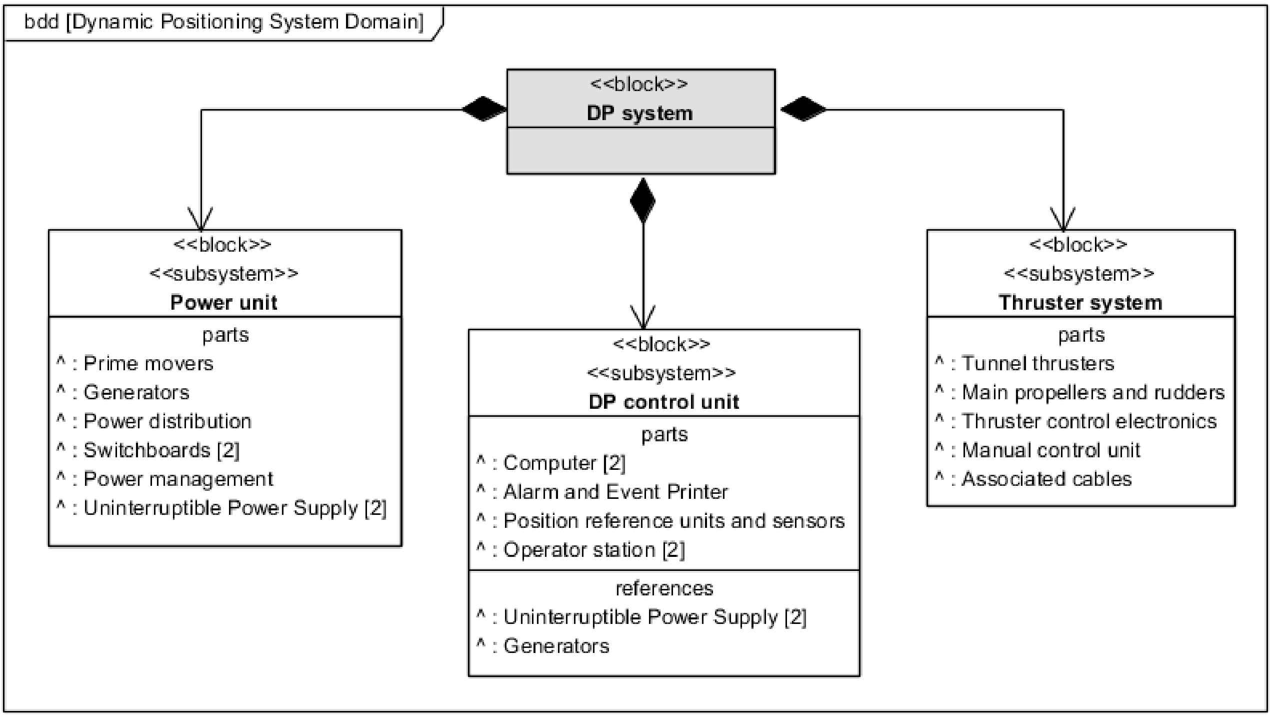


Figure 3: The block definition diagram presenting the structure of DP-system domain

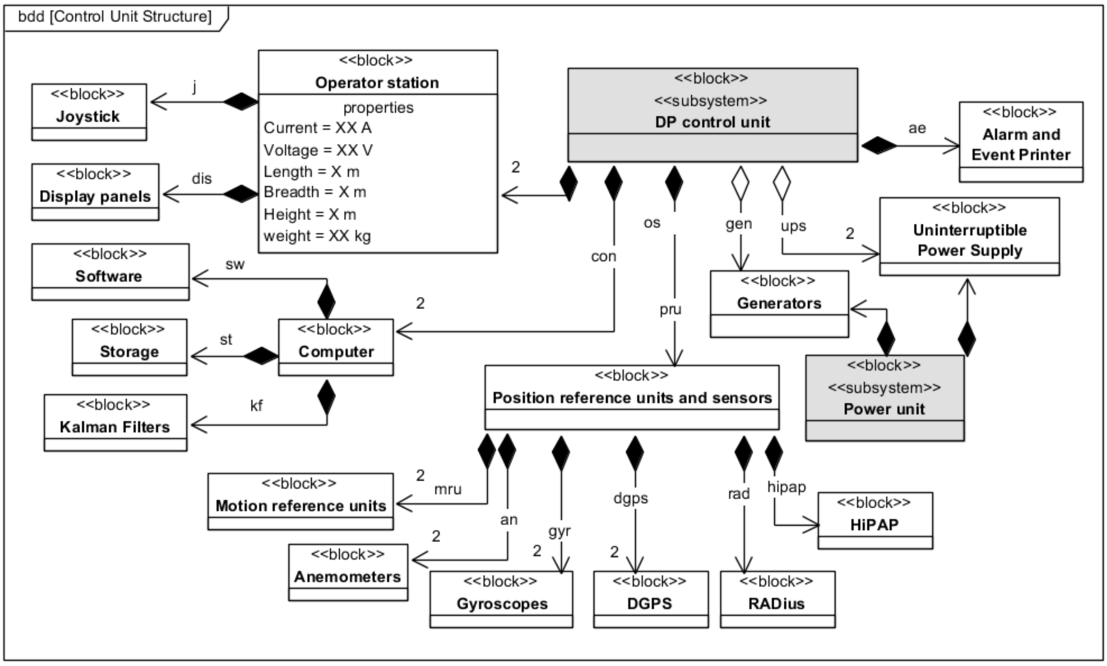


Figure 4: The block definition diagram presenting the structure of the control unit

Figures 5 and 6 present the parametric diagram for analyzing power consumption of control unit and power efficiency of the DP system respectively. This is achieved by importing the property values (see the Operator station block in Figure 4) and using power equations.

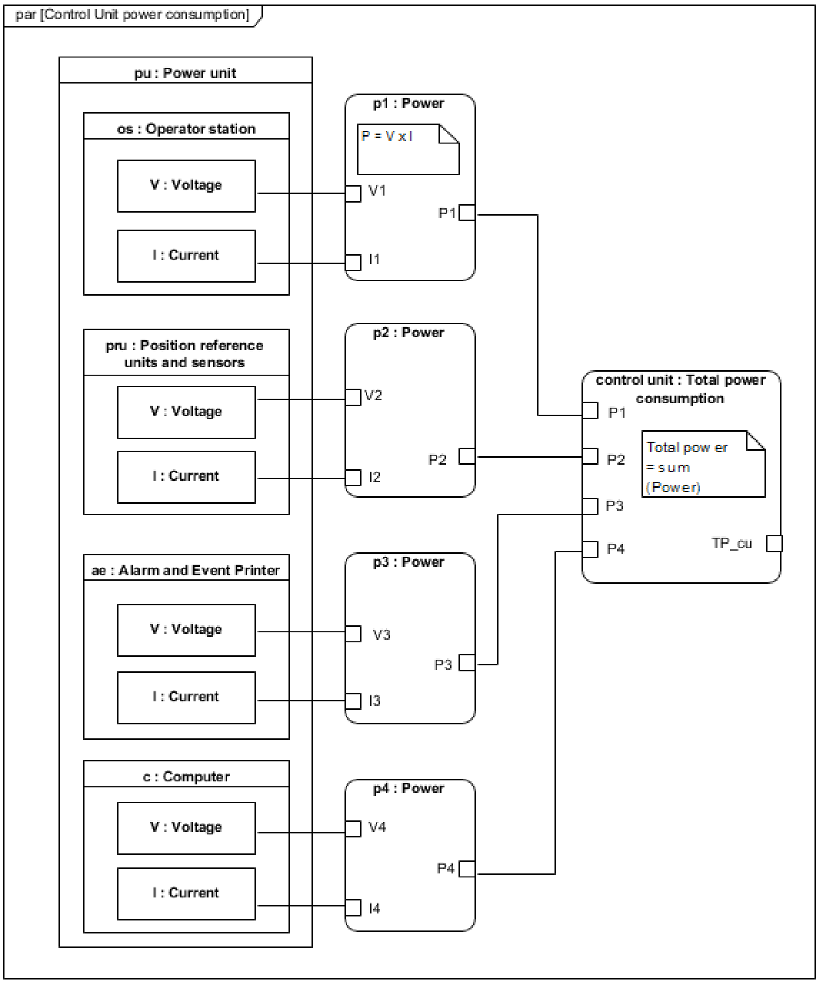


Figure 5: The parametric diagram for calculating power consumption of DP control unit

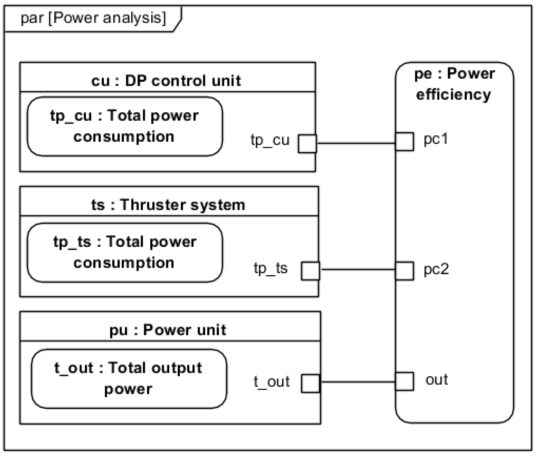


Figure 6: The parametric diagram for power analysis

After modeling the structural aspects of the system, the diagrams presenting the behavioural aspects are created. Figure 7 presents the use cases and the users that are involved in automatic vessel positioining. The <<include>> stereotype denotes the cases that are always active during the main operation, while <<Extend>> stereotype denotes the cases that are only active in certain scenarios.

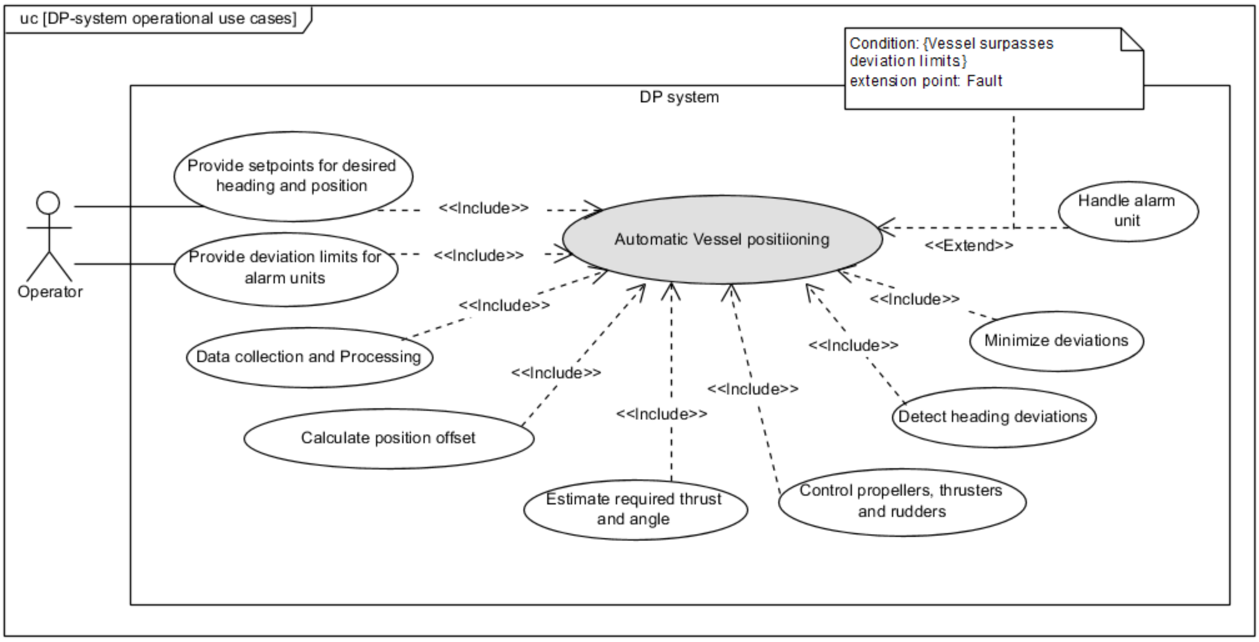

uc [DP-system operational use cases]
DP system
Condition: {Vessel surpasses deviation limits}
extension point: Fault
Operator
Provide setpoints for desired heading and position
Provide deviation limits for alarm units
Data collection and Processing
Calculate position offset
Estimate required thrust and angle
Control propellers, thrusters and rudders
Detect heading deviations
Minimize deviations
Handle alarm unit
Automatic Vessel positiioning
<<Include>>
<<Extend>>

Figure 7: The use case diagram presenting use cases of DP-system

Figures 8 and 9 present the activity diagrams for the DP operation and its sub-function (data collection and filtering) respectively. It presents the sequence of the tasks involved in the process using direction of arrows. The tasks in parallel are shown with Join nodes and Fork nodes (black filled rectangles). In addition to the sequence of the tasks, the activity diagram of data collection and filtering process presents the involved components, required inputs (satellite, transponders etc.) and outputs (filtered positional data, wind speed data etc.).

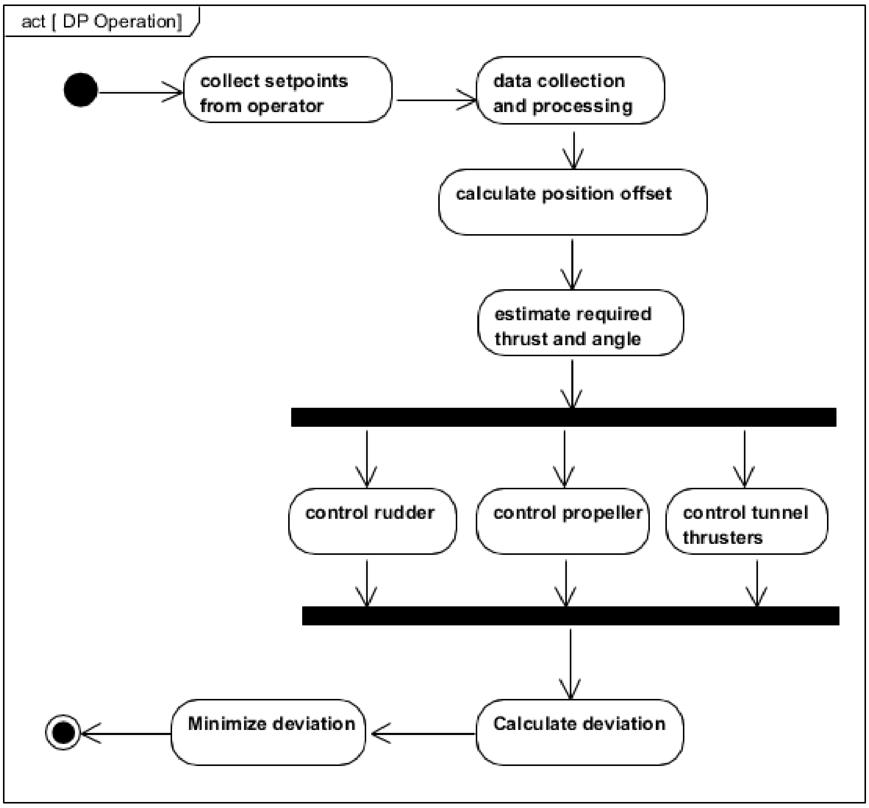

act [DP Operation]
collect setpoints from operator
data collection and processing
calculate position offset
estimate required thrust and angle
control rudder
control propeller
control tunnel thrusters
Calculate deviation
Minimize deviation

Figure 8: The activity diagram presenting the actions during DP operation

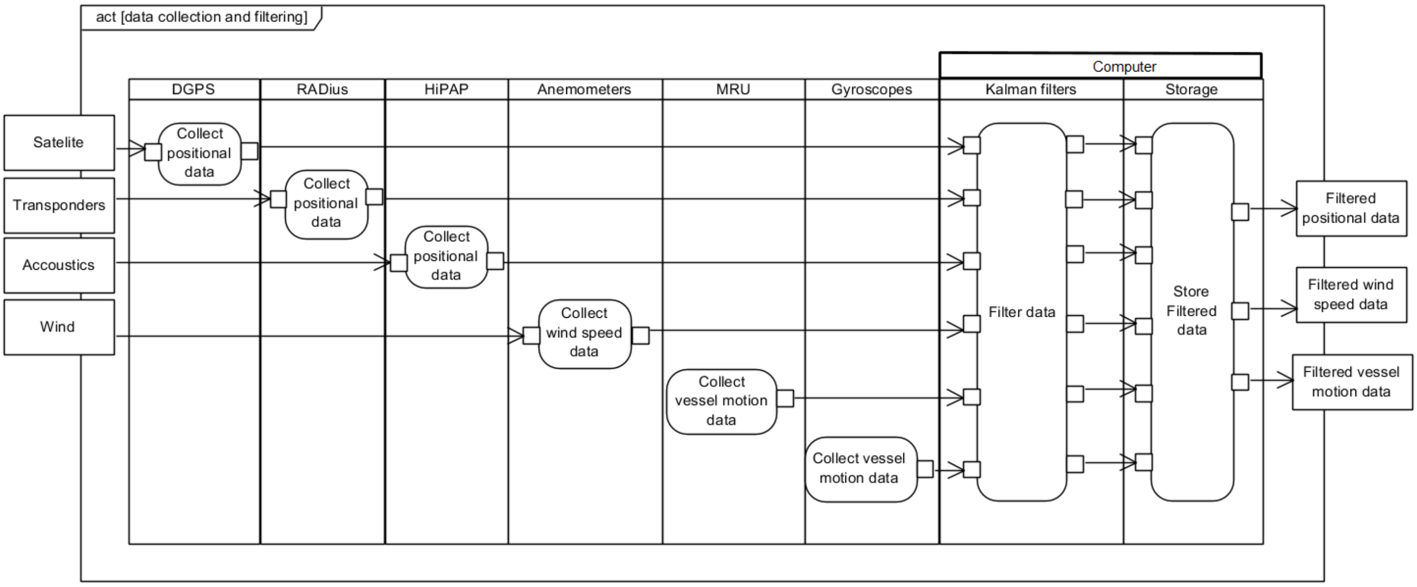


Figure 9: The activity diagram for data collection and filtering process

Figure 10 presents the state machine diagram for "handle alarm unit" use case. The diagram presents different states of the alarm unit (Alarm off, Alarm on etc.), their triggering causes (e.g. Activate warning if..), system functions during that specific state (e.g. Activate warning message and signals). All of these behavioural diagrams are then placed in their corresponding packages in the package diagram.

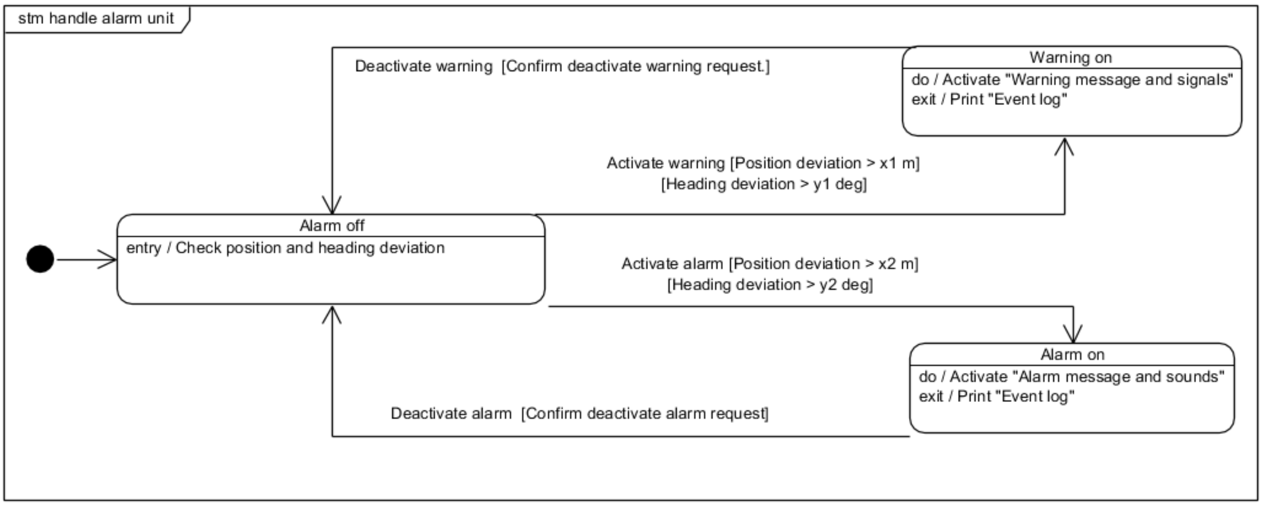

stm handle alarm unit
Deactivate warning [Confirm deactivate warning request.]
Warning on
do / Activate "Warning message and signals"
exit / Print "Event log"
Activate warning [Position deviation > x1 m]
[Heading deviation > y1 deg]
Alarm off
entry / Check position and heading deviation
Activate alarm [Position deviation > x2 m]
[Heading deviation > y2 deg]
Alarm on
do / Activate "Alarm message and sounds"
exit / Print "Event log"
Deactivate alarm [Confirm deactivate alarm request]

Figure 10: The state machine diagram for handling alarm unit of DP-system

3.3 Modeling with OPM using OPCAT tool

Figure 11 presents the System Diagram (SD, top hierarchical level) for Automatic vessel positioning. It shows the main process (automatic vessel positioning), involved objects (e.g. operators, individual DP-systems, etc.) and their links. Furthermore, it also presents the state transition of an object (vessel) from initial position to required position achieved by automatic vessel positioning function.

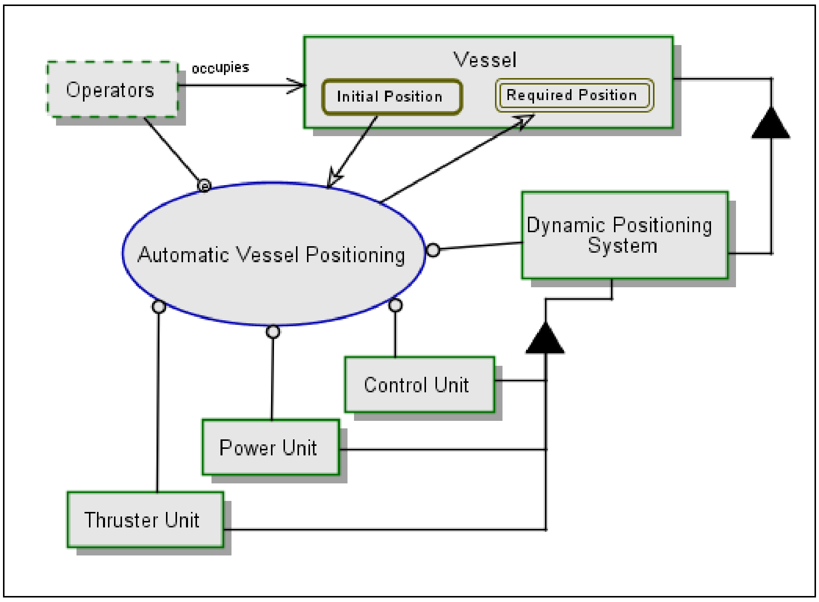


Figure 11: The top level OPD (SD) for DP-system operation

Figure 12 presents the In-Zoomed diagram of the SD. It shows the subprocesses and involved components in the automatic vessel positioning process. The sequence of the subprocesses is represented using a top-down approach.

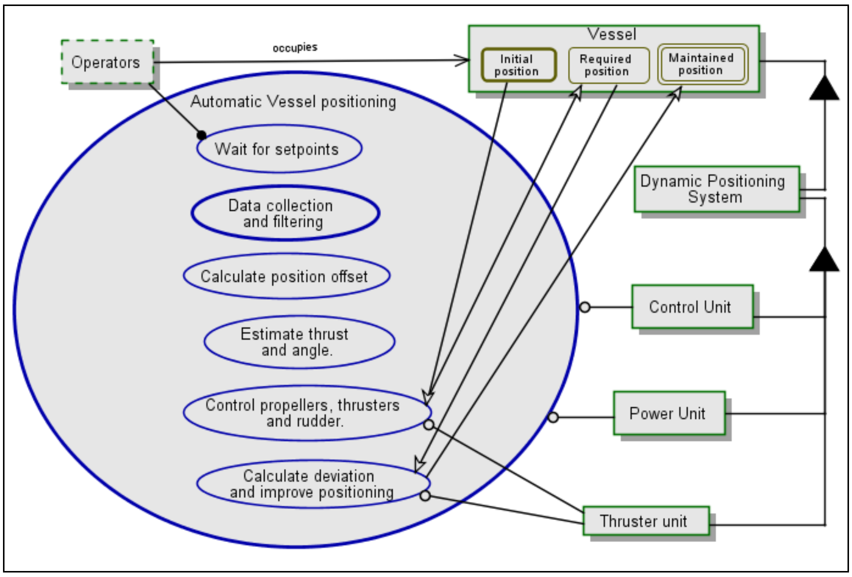

Operators
occupies
Vessel
Initial position
Required position
Maintained position
Automatic Vessel positioning
Wait for setpoints
Data collection and filtering
Calculate position offset
Estimate thrust and angle.
Control propellers, thrusters and rudder.
Calculate deviation and improve positioning
Dynamic Positioning System
Control Unit
Power Unit
Thruster unit

Figure 12: The In-Zoomed diagram of automatic vessel positioning

The OPD is then further refined by adding In-zoomed diagram for each of the subprocesses. Figure 13 presents the In-zoomed diagram for the subprocess “Data collection and filtering”. It presents the subprocesses and objects involved during Data collection and filtering.

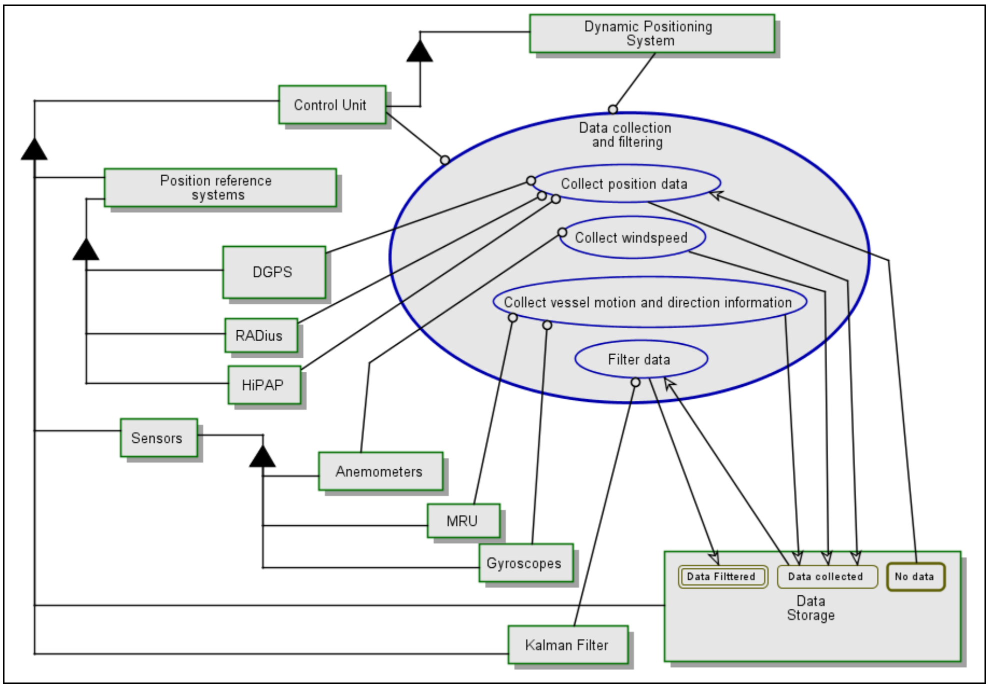


Figure 13: The Zoomed-in diagram for data collection and processing

The OPL of each of these diagrams were auto-generated. The OPL for SD is as following:

Vessel is physical.
Vessel can be Dynamic or Maintained position.
Dynamic is initial.
Maintained position is final.
Vessel consists of Dynamic Positioning System, Power Unit, and Thruster unit.
Dynamic Positioning System is physical.
Dynamic Positioning System consists of Control Unit.
Control Unit is physical.
Power Unit is physical.
Thruster unit is physical.
Operator is physical.
Operator triggers Vessel positioning.
Vessel positioning requires Control Unit, Operator, Dynamic Positioning System, Thruster unit, and Power Unit.
Vessel positioning changes Vessel from Dynamic to Maintained position.

Figure 14 presents the simulation of the SD diagram of DP-System where the left diagram represents the early stages and the right diagram represents the final stages . The simulation demonstrates the change of states and the sequence of the processes in real-time. The green color in the simulation refers to the active objects and the purple colour refers to the active processes. Furthermore, the change of state of an object is demonstrated with a moving red circle in the links and the changing color of the state itself.

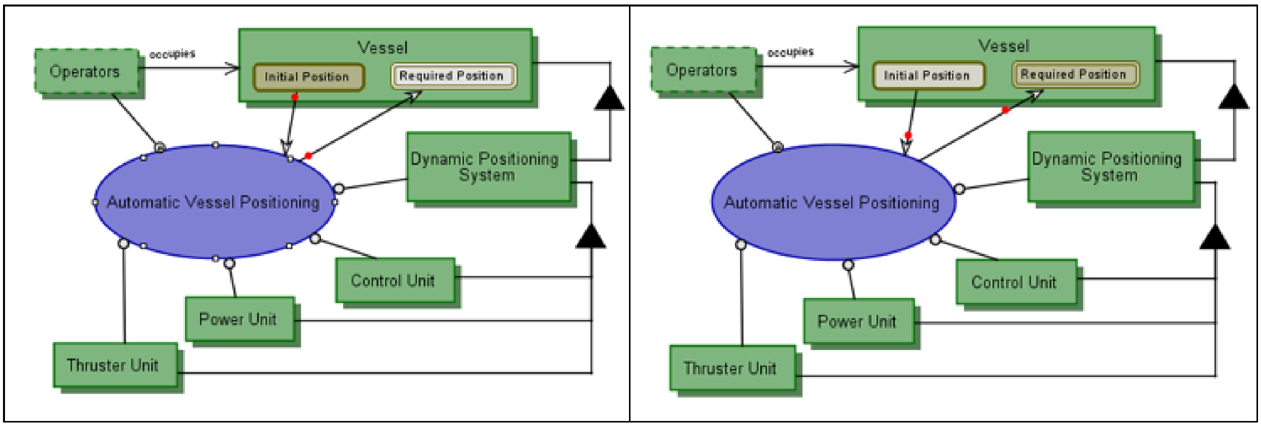


Figure 14: A simulation presenting the state change of the vessel (early stage of simulation at left and final stage of simulation at right)

4 DISCUSSION

4.1 Comparison of OPM and SysML

Based on the generated models, Table 1 summarizes the key aspects and strengths of each methods.

Table1: A summary of key aspects of OPM and SysML

Features	SysML	OPM
Modeling structural and behavioral aspects	• Detailed information of the system using different type of diagrams. • Distinction between system owned and shared components.	• Distinction between physical and non-physical objects. • Easier and faster to create and understand as it consists a single unified diagram.
Models management	• An entire diagram "package diagram" is dedicated for the model's management.	• Zoom-in and zoom-out approach is used to easily navigate between the levels. • Easier to manage because it consists of a single diagram (9 in SysML)
Availability of tools support and related features	• A wide range of tools is available in the market with limited or full features of SysML.	• The only available tool offers full features of OPM
Additional features	• Includes Requirement diagram that presents the requirements of the system elements and allows engineers to verify and validate these requirements during system life cycle. • Includes Parametric diagram that enables various engineering analysis such as trade studies, sensitivity analysis, performance analysis and design optimization.	• Allows dynamic simulation of the model, which enables the modeler to visualize the system functions and test its executability. • Generates texts describing the models automatically which can be useful for technical and non-technical stakeholders. • Allows coders to make models using codes instead of using graphical user interface.

The results show that both methods are effective in modeling the structural and behavioral aspects of the system. However, SysML provides more detailed information of the system than OPM as it consists of several types of diagram. In contrast, OPM uses a single unified diagram to present the system composition and behavior in a simple way. Due to these differences, OPM models are easier and faster to create and understand than SysML models.

There are also some important unique features in both methods. OPM allows dynamic simulation of the system. Whereas in SysML, the simulation is only possible for certain diagrams and with specific tools. Other key features available in OPM is the ability of textual modality (automatic generation of texts from the models) and also allows coders to make models using codes. On the counterpart, SysML includes requirement diagram and parametric diagram which are lacking in OPM. The requirements diagram in SysML can include the requirements that must be satisfied during the design, development and even during the operation of the system. The parametric diagram enables to conduct various engineering analysis of the systems such as trade studies, sensitivity analysis, performance analysis and design optimization.

4.2 Applicability of methods to autonomous ship systems

It is highly expected that the autonomous ship systems will consist of tightly coupled components with complex interactions within the system and with the environment (Levander, 2017). The functions of operators onboard will be added to the system components (i.e. software and hardware), which are already operating with many complex functions. The functions that cannot be assigned to the existing system components will be assigned to the new systems with advanced components. As a result, it is evident that an autonomous ship will include high number of advanced systems on board with higher interactions between them when compared to the traditional ships. Thus, the current approach, where the system description is mostly attempted to communicate with texts and traditional approaches such as tree structure and functional block diagrams might not be enough to understand these complex interactions and to communicate them. Furthermore, the advancement of information technology, and the availability of standard methods and tools allows the engineers to move from the document-centric approach to model-based methods.

Both OPM and SysML consist of features that can be useful for the development, analysis and operation of autonomous ship systems. However, there are some important features such as requirements and parametric modeling that are lacking in OPM. Marine industry is highly dependent on rules and regulations. There are several requirements such as DNV GL (2013) from regulatory bodies that need to be fulfilled before ships deployment. Thus, it is necessary to model the requirements of ship systems and their operation from the earliest design phase to avoid any potential issues afterwards. Furthermore, the parametric diagram in SysML allows engineering analysis to be conducted from the design phase and throughout the system lifecycle. Although, there has been an attempt to include requirements and parametric in OPM models (Dori, 2016), the addition of requirements and parameters in an OPD diagram may lead to information overload. Furthermore, as autonomous ships will consist of high number of components and interactions, the attempt of adding all information in a single diagram will make models large and complex to understand. Although, the features of OPM such as object process language and simulation can prove beneficial for engineers and different stakeholders, the features available in SysML are of higher importance. Nonetheless, it is still necessary to assess the suitability of adding these missing features in SysML, which can make SysML even more complete and applicable to autonomous ship systems.

Ships in general are safety-critical and the development of autonomous ships are of high interest among several stakeholders in marine industry. Thus, it is expected that the models should be as detailed as possible. As SysML provides modeling capabilities for systems in higher details than OPM and considering all of above reasons, SysML can be more suitable than OPM for modeling the Autonomous ship systems.

4.3 Exploring the model’s utilization for analyzing autonomous ship systems.

Models can provide clear and attractive communication of a system’s specification and the interaction of systems in a module. Models using MBSE methods have been increasingly used in several engineering analysis as presented in Section 1.1. The following topics related to autonomous ship systems will be explored in the future for assessing the possibility of integrating models into the methods:

- **Hazard analysis**: For identifying and analyzing the hazards in the system, the analysts must understand the system composition and behavior. For this purpose, the system information generated by the models can be used to understand the autonomous ship systems before conducting hazard analysis. Instead of relying heavily on experts and their brainstorming sessions, these models can be potentially used as an input for experts. Furthermore, the experts who are less familiar with the system can use the models to gather necessary system information required for identifying hazards. The challenge is then to identify the most important diagrams to be communicated to the analysts as providing everything can lead to the problem of information overload.

- **Real-time system monitoring**: There is also a possibility to use the models as an interface for real-time system monitoring for autonomous ships. A simple example would be by adding a component status property to each of the components block in SysML where they are linked to the sensors installed on the ship. If the sensor is not sending any signals, the model will notify the operator by changing the status from “operating” to “non-operating”. In addition, it will display all other components and activities that are affected by this status change and the safety measures that need to be followed as predefined by the system’s designers. Thus, the possibility of using SysML diagrams for real-time system monitoring of autonomous ship systems should be further explored in future.

- **System maintenance:** Similar to component status, the information about the expected lifetime of each component and the guidelines for executing the maintenance from the manufacturers can be also added to the diagrams and communicated to the maintenance engineer. Furthermore, a property can be added in the blocks of the components which records the date of previous maintenance. This will allow engineers to keep track of the components and use the information stored in models to guide the maintenance process.

Most of the engineering or operating tasks that require detailed information about the system could be guided by these system models. However, the implementation of these methods is heavily reliant on the tools. Thus, the tools that are currently being used need to be updated and assured for conducting these analyses for autonomous ship systems.

5 CONCLUSIONS

Graphical models could unlock various ways of communicating system information to relevant stakeholders. Thus, system modeling methods could be a viable option for autonomous ship systems. This paper explored OPM and SysML for handling the key issues related to complex systems i.e. autonomous ship systems. These key issues are complexity management, system understandability and communication of the system information. The methods were applied for the case of a typical Class II DP-system, that is believed to have a major role in autonomous ship operations. The in-depth comparison shows that both methods have been able to manage the system complexity and communication of system information using graphical illustrations or models.

Although, the complexity management and communication are covered well by both methods, SysML is more effective than OPM in handling the system understandability. As SysML provides different diagrams that provide system information from different perspective, it is easier to understand the system structure and behavior in all system hierarchical level. Furthermore, SysML consists of requirements diagrams and parametric diagram for supporting the requirements analysis and other engineering analysis (i.e. sensitivity analysis, performance analysis and design optimization), which is lacking in OPM. Thus, SysML can be more suitable than OPM for modeling the autonomous ship systems. However, OPM models are easier and faster to create and understand than SysML and should be implemented in those cases with limited resource availability and where the level of detailed information provided by OPM models are sufficient for the analysts. Furthermore, adding the distinct features of OPM such as simulation and automatic text generation to the SysML should be explored in the future.

The discussions and conclusions of this research are based on the comparison of DP-system models generated using the system operation manual. For improving the creditability for the research, these models and the results will be assessed using expert’s opinion in the future.

ACKNOWLEDGEMENTS

This research has been funded by RESET project. RESET has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 730888.

REFERENCES

Best, B., Jurjens, J., & Nuseibeh, B. (2007). Model-Based Security Engineering of Distributed Information Systems Using UMLsec. *29th International Conference on Software Engineering (ICSE'07)*, (pp. 581-590). Minneapolis. doi: 10.1109/ICSE.2007.55

D'Ambrosio, J., & Soremekun, G. (2017). Systems engineering challenges and MBSE opportunities for automotive system design. *International Conference on Systems, Man, and Cybernetics*, (pp. 2075-2080). Banff.

DNV GL (2013). Rules for Classification and Construction. Germanischer Lloyd SE, Hamburg. Retreived from: http://rules.dnvgl.com/docs/pdf/gl/maritimerules/gl_i-1-15_e.pdf

Dori, D. (1995). Object-process Analysis: Maintaining the Balance Between System Structure and Behaviour. Journal of Logic and Computation, 5(2), 227-249. Available at:

https://doi.org/10.1093/logcom/5.2.227Dori, D. (2002). *Object-Process Methodology- A Holistic Systems Paradigm.* Berlin: Springer-Verlag .

Dori, D. (2016). *Model-Based Systems Engineering with OPM and SysML.* Cambridge: Springer.

Enterprise Systems Modeling Laboratory. OPCAT - OPM tool. Available at: http://esml.iem.technion.ac.il/

Friedenthal, S., Griego, R., & Mark, S. (2007). INCOSE Model Based Systems Engineering (MBSE) Initiative. *INCOSE .* San Diego.

Friedenthal, S., Moore, A., & Steiner, R. (2015). Systems Engineering Overview. In *Practical Guide to SysML - The systems Modeling Language*.

Grobshtein, Y., Perelman, V., Safra, E., & Dori, D. (2007). Systems Modeling Languages: OPM Versus SysML. *International Conference on Systems Engineering and Modeling* , (pp. 102-109). Haifa. doi: 10.1109/ICSEM.2007.373339

Hollnagel, E. (2012). *FRAM: The Functional Resonance Analysis Method.* Ashgate Publishing Limited.

Holt, J., & Perry, S. (2019). In *SysML for Systems Engineering - A Model-Based Approach (3rd Edition)* (p. 3). London: The Institution of Engineering and Technology.

International Maritime Organization. (1994). *Guidelines for vessels with Dynamic Positioning Systems.* Retrieved from http://www.imo.org/blast/blastDataHelper.asp?data_id=10015&filename=MSCcirc645.pdf

International Standards Organization. (2015). *ISO/PAS 19450:2015 - Automation Systems and Integration -- Object-Process Methodology.* Retrieved from https://www.iso.org/standard/62274.html

International Standards Organization. (2017). *ISO/IEC 19514:2017- Informational Technology -- Object Management Group Systems Modelling Languuage (OMG SysML).* Retrieved from https://www.iso.org/standard/65231.html

Kongsberg (2014). *Kongsberg K-Pos DP (OS) Dynamic positioning system with Offshore Loading Application.* Kongsberg Maritime AS.

Levander, O. (2017). *Autonomous ships on the high seas. IEEE, 54*(2), 26-31.

Leveson, N. G., & Thomas, J. P. (2018). *STPA HANDBOOK.* Retrieved from https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf

Mhenni, F.¨, Nguyen, N., Kadima, H., & Choley, J.-Y. (2013). Safety Analysis Integration in a SysML-Based Complex System Design Process. *Systems Conference (SysCon).*

MUNIN. (2016). *Research in Maritime Autonomous Systems Project Results and Technology Potentials.* Retrieved from, http://www.unmanned-ship.org/munin/wp-content/uploads/2016/02/MUNIN-final-brochure.pdf

Object Management Group. (2017). OMG Systems Modeling Language. Retrieved from https://sysml.org/.res/docs/specs/OMGSysML-v1.5-17-05-01.pdf

Russell, M. (2012). Using MBSE to Enhance System Design Decision Making. *Procedia Computer Science* (pp. 188-193). Elsevier.

Visual Paradigm 15.2. SysML tool. Available at: https://www.visual-paradigm.com/

Weck, D., Roos, O. L., Magee, D., & L., C. (2011). Modeling and Analyzing Engineering Systems. In *Engineering system - Meeting Human Needs in a Complex Technological World.* MIT Press.

An initial hierarchical systems structure for systemic hazard analysis of autonomous ships

Meriam Chaal[*], **Osiris Valdez Banda**, **Sunil Basnet**, **Spyros Hirdaris** and **Pentti Kujala**

Aalto University, Department of Mechanical Engineering, Marine Technology, Espoo, Finland

Abstract

Safety assurance of autonomous ships is one of the major long-term challenges faced by the maritime world. Applying systemic hazard analysis methods at this early stage will guide the design and operation of safe autonomous ships. This paper proposes an initial hierarchical ship systems structure that could be the basis for a systemic hazard analysis of autonomous ship systems and operations. The approach is based on the systems theory and the principle of hierarchy and has been developed via the combination of models used in past research projects and requirements of the STCW convention. For enabling the operation of autonomous ships, the ship crew functions are either replaced by ship technical systems or assigned to the Shore-Based Control Centre (SCC).

Keywords: Autonomous ship systems; Autonomous Navigation System; Situation awareness, systems theory, system of systems, system function, systemic hazard analysis

---

[*] Corresponding author: +358 50 432 0739 meriam.chaal@aalto.fi

1. Introduction

Autonomous ships aim at improving the safety and efficiency of the maritime operations while also preventing the exposure of the ship crew to on-board hazards (Wróbel et al., 2017). The specification of requirements and procedures for safety assurance in autonomous ships is complex and risks must be accounted for at early design stage. This challenge is also reflected in International Maritime Organization (IMO) who require that future Maritime Autonomous Surface Ships (MASS) should operate at an equivalent level of safety (i.e. be ''at least as safe as'') conventional vessels (IMO, 2018).

Autonomous ships are expected to be highly complex with software-intensive interacting systems, that require the application of systemic hazard analysis methods that capture hazardous systems interactions (Basnet et.al , 2019). These methods assume that the ship is a system comprising of sub-systems that interact with each other (Leveson, 2011; Valdez Banda et al., 2019).To conduct this hazard analysis, a hierarchical systems description of the autonomous ship is necessary.

In an attempt to open the way toward such developments, this paper reviews results of two of the major research projects in the field of autonomous ship operations namely MUNIN (Maritime Unmanned Navigation through Intelligence in Networks) and AAWA (Advanced Autonomous Waterborne Applications). Consequently, based on lessons learnt, systems theory and STCW functional requirements it suggests an initial hierarchical systems structure for the risk assessment of an autonomous ship at a level of autonomy AL4. In this sense, it contributes toward developing a new framework for hazard analysis beyond classical methods such as the IMO classic Fault Tree Analysis currently used for the development of rules and regulations within the context of Formal Safety Assessment.

2. Necessity of systemic hazard analysis for autonomous ship systems

2.1. Systemic hazard analysis

Systems theory was introduced in 1930's to cope with the complexity of the systems starting to be built in different domains at that time (Ackoff, 1971; Leveson, 2011). The approach defines complex systems as systems of systems, where every system has a function (or purpose), elements (or components), and interconnections (Arnold & Wade, 2015). According to the hierarchy principle in the systems theory, each system at its level could be a sub-system at a higher level and a set of sub-systems at a lower level (Adams, 2011). The sub-systems interact and work together to perform their main system function and cannot be decomposed into independent physical components (Adams, 2011).

Systems thinking applied to safety revealed that safety is a system property that is affected by the interactions of its components (Hollnagel, 2004; Leveson, 2004). The hazards emerging from these interactions lead to unexpected accidents that were not considered in the traditional risk assessments (Hollnagel, 2004; Leveson, 2004). The systemic hazard analysis methods came as a response to the limitations of the traditional hazard analysis and risk assessment techniques in identifying the hazards associated with the interactions (Aven, 2016; Leveson, 2011).

Numerous traditional linear (cause-effect) hazard analysis methods have been developed and applied to different systems that humans had designed. The most widely applied are Fault Tree Analysis, Event Tree Analysis and HAZOP, which were developed many decades ago. These methods were successful in hazard analysis and risk assessment of relatively simple technical systems (Altabbakh, AlKazimi, Murray, & Grantham, 2014). However, the same techniques applied to the modern complex sociotechnical systems have shown very less effectiveness as they focus only on the components' failure in a linear causal analysis, which cannot detect the non-linearity in today's complex systems (Aven, 2016). In addition, these methods rely on historical data of the system, which puts the risk decisions under an increased uncertainty about the knowledge of the emerging technologies that do not have historical data (Aven, 2016; SRA, 2015). Therefore, modern complex systems need a systemic approach for hazard analysis and risk assessment in order to consider both the hazards related to components failures and the new hazards emerging with components interactions.

The most popular system theoretic hazard analysis methods as employed in the literature are STPA (System Theoretic Process Analysis) and FRAM (Functional Resonance Analysis Method). FRAM and STPA applied to complex modern systems have been successful in coping with complexity of the modern systems and capturing the hazards associated with their components interactions (Patriarca et al., 2017; Valdez Banda and Goerlandt, 2018).

2.2. Autonomous ships as complex systems

Autonomous ships are systems with embedded software and high functional dependencies and integration. This makes them complex systems, where a software may control separated subsystems, and depend on other systems operating across the physical boundaries (Utne et al., 2017)

As explained in the previous section, systemic hazard analysis could then be applied to autonomous ships as complex systems and prevent the hazardous scenarios related to both their components and interactions. Applying these modern techniques at the systems development and design stages could improve safety (Fleming et. al, 2013; Ishimatsu et al., 2014; Valdez Banda et al., 2019). The results of the systemic hazard analysis of autonomous ships will then contribute to their safe deployment. The representation of the autonomous ship as a system of systems working together to perform the autonomous ship function would allow the systemic hazard analysis at this early stage of its development.

2.3. Autonomous ship functions and the role of humans in the loop

Fully autonomous ships are supposed to perform all previous functions of the technical systems and hence compensate the human absence. In addition, autonomous ships with their different levels of autonomy should be at least as safe as conventional ships as prescribed by the IMO (IMO, 2018).

The lack of experience in designing and operating autonomous ships justifies the need to employ the experience gained in designing and operating traditional ships. Besides, "*the autonomous ships will most likely remain ships*" and will navigate and behave like conventional ships (Wróbel and Montewka, 2019), which justifies more the need to consider the experience gained in conventional ship operations. Furthermore, the development of the autonomous systems started already by replacing the human capabilities when developers identified the required technologies to replace the human senses during navigation. Some of the suggested technologies were for example cameras and microphone arrays to compensate the human visual and hearing capabilities respectively.

The IMO standards have been continuously amended to hold the experience gained through the design and operation of conventional ships. The International Convention on Standards of Training, Certification and Watch-keeping for Seafarers (STCW) is one of the main IMO legal instruments that has continuously accommodated the updates in the functions of the ship crew based on the experience gained in the ship operation.

3. Review of autonomous ship technical concepts

3.1. MUNIN

The European project "Maritime Unmanned Navigation through Intelligence in Networks" (MUNIN) was the first research project dedicated on developing the technical concept of an autonomous cargo ship. It studied the feasibility and safe implementation of the concept with tests on an existing dry bulk carrier (MUNIN, 2016). The concept suggested that for a simple first application and with the limitation of the connectivity bandwidth, the autonomous ship should be able to sail in open seas most of the time under full autonomy mode (MUNIN, 2015). The definition of the concept has been supported by different IMO conventions including the STCW convention. As shown in Figure 1, five new systems namely : (a) an Advanced Sensor Module (ASM), (b) an Autonomous Navigation System (ANS), (c) an Autonomous Engine Monitoring and Control System (AEMCS), (d) an Autonomous Ship Controller (ASC) and (e) a Shore Control Centre (SCC) were suggested to be essential for the safe operation of autonomous ships in deep sea. In addition, the two old Bridge Automation System and Engine Automation System were existing in the use case ship. The port approaches and special manoeuvres were excluded from the autonomous operation in order to reduce complexity for the early applications.

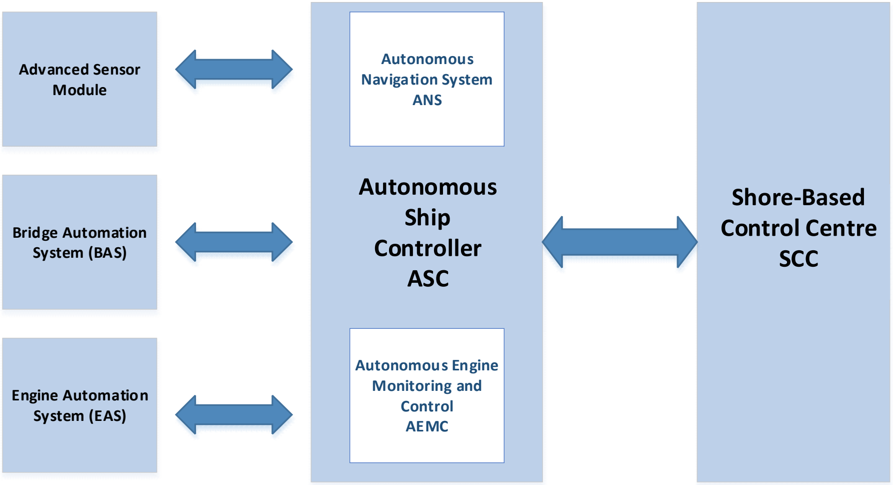

Advanced Sensor Module
Bridge Automation System (BAS)
Engine Automation System (EAS)
Autonomous Navigation System ANS
Autonomous Ship Controller ASC
Autonomous Engine Monitoring and Control AEMC
Shore-Based Control Centre SCC

Figure 1: Overview of the autonomous ship modules (MUNIN, 2013)

The advanced sensor module was created in order to complement the absence of humans on-board in performing the lookout function (C. Bruhn, Burmeister, T. Long, & A. Moræus, 2014). The ANS function is “navigating the unmanned autonomous ship safely from boarding point to boarding point (MUNIN, 2015). Under this main function, the ANS should conduct weather routing, determine ship dynamics, control buoyancy and stability, avoid collision and manage alarm and emergencies (MUNIN, 2015). The AEMCS monitors and controls all the engine room systems. The ASC assesses the data from different ship sensors and from the shore and controls the autonomous ship operation. The SCC conducts the voyage planning with the administrative tasks, manages the distress communication and monitors the overall ship operation to manage the complex emergencies. The BAS and the EAS would have to perform the same functionalities in the existent ship; the BAS receives navigation alerts through NAVTEX, keeps log book and follows track with autopilot, while the EAS provides engine data.

The project results recognised that satellite bandwidth and communication quality are great challenges to a full-time remote operation and argued that the ship should be able to operate autonomously most of the time. The same challenge was also recognized by AAWA project late and a dynamic level of autonomy during the voyage was suggested (AAWA, 2016). Thus, the SCC will serve as back up with a remote control in special manoeuvres and critical situations (Rødseth et al., 2013). One more backup system is the “fail to safe” situation, when both the autonomous ship controller and the SCC fail to control the ship or execute the adequate tasks. In this emergency case, the ship should follow a predefined set of actions or route that takes it to a safe situation without considering its initial plan execution.

3.2. AAWA

In AAWA project, the leading Rolls Royce marine group with other partners from the maritime industry and academia sought for applicable technologies with site tests in a defined testbed in Finland. The focus was on autonomous navigation systems. The project defined the autonomous navigation architecture as a set of modules that could together enable the safe navigation of the autonomous ship from port A to port B (AAWA, 2016). Figure 2 illustrates four modules of the ANS (Ship State Definition, Route Planning, Collision Avoidance, Situational Awareness and Dynamic Positioning).

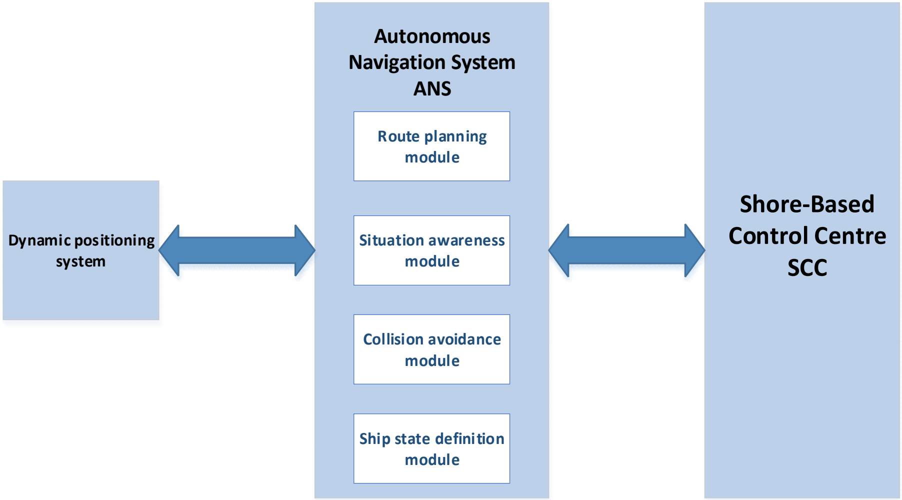

Autonomous Navigation System ANS
Route planning module
Situation awareness module
Collision avoidance module
Ship state definition module
Dynamic positioning system
Shore-Based Control Centre SCC

Figure 2: Autonomous Navigation System (ANS) architecture, (AAWA, 2016)

The Ship State Definition module also known as “Virtual Captain” is the highest in the hierarchy of the ANS architecture and it receives and processes information from the other modules in order to make decisions based on a full awareness of the ship situation.

The Route planning module on the other hand is charged of delivering the route plan based on a software that considers information provided in the voyage plan received from the shore. It generates the route on a static mode, while the Collision Avoidance module generates the dynamic path to avoid collision during the plan execution.

The Situation Awareness module fuses the data from its different sensor types and extracts the adequate information to map the ship surroundings. The surroundings map is necessary for the Collision Avoidance module. The project suggested that the currently available sensors technologies can provide the lookout required for a safe navigation if the adequate sensor fusion combinations for each situation are determined (AAWA, 2016). The identified set of sensors having the potential to replace the human lookout were HD cameras, IR cameras, Radar, short range Radar, Lidar (Light Detection and Ranging) and microphones.

The Dynamic Positioning (DP) module controls the propulsion system in order to track a defined route or keep a defined position. Furthermore, the DP module monitors the ship motion and manoeuvrability constraints in order to act accordingly.

4. Combining MUNIN and AAWA concepts with reference from STCW functions

In order to conduct future systemic analysis that are aligned with the development trends, the contribution of MUNIN and AAWA projects should be considered in describing the hierarchical ship systems structure. For this reason in this paper, the combination of the technical concepts in MUNIN and AAWA is considered as the starting point to develop the autonomous ship systems structure. A cross verification with the systems identified based on the seafarers’ functions in STCW convention is then conducted to add the missing functions in the structure or merge the systems with same functions. First, the general concept of MUNIN application is employed because it was not restricted to the ANS and it gave the context of operation, which could be considered as a level of autonomy AL4 in Lloyds Register’s definition. Then, as AAWA project was focused mainly on the ANS and the research was conducted more recently, and gave more details about the system’s technologies to be employed, it was combined with the general concept of MUNIN.

Table 1 illustrates the transition from the different functions of the STCW convention to the correspondent autonomous ship systems that would perform these functions. As each system should have a purpose and a set of sub-systems, each function in the STCW convention at the operational level could be assigned as the purpose of a system in the autonomous ship.

Table 1: STCW functions and the correspondent autonomous ship systems

STCW functions			Autonomous ship systems			
			MUNIN	**AAWA**	**Other ship systems/sub-systems**	
Deck department	**Navigation**	Plan route		Route planning system (Sub)		**ANS**
		Determine position speed and heading			Positioning, Navigation and Timing System	
		Reporting			Communication and reporting system	
		Track route		DP system (Sub)		
		Interpret and apply weather information			Weather monitoring and interpretation system	
		Avoid collisions (consider COLREGs lights and sound signals)		Collision avoidance system (Sub)		
		Maintain proper lookout		Navigational situation awareness system (Sub)		
		Respond to emergencies	SCC			
		Respond to distress signals	SCC			
		Maneuvering		DP system (Sub)		
		Anchoring and mooring			**Anchoring and Mooring system**	
	Cargo handling and stowage	Loading and stowage			**Cargo handling and monitoring system**	
		Unloading				
		Care during the voyage				
	Controlling the operation of the ship and care for persons on board	Compliance with pollution prevention	SCC			
		Ship stability (Trim, buoyancy, watertight integrity)			**Ship stability and integrity system**	
		Fire prevention, control and fighting				
		Life-saving appliances operation	Not applicable to Cargo ships			
		Medical aids application	Not applicable to Cargo ships			
		Compliance with legislation	SCC			
Machinery department	**Marine engineering**	Use hand tools, electrical and electronic measurement and test equipment for fault finding, maintenance and repair operations			**Maintenance planning system**	
		Monitoring of engineering equipment and systems	**AEMC system**			
		Apply safety and emergency procedures for all systems	**AEMC system**			
		Operate main and auxiliary machinery	**AEMC system**		Propulsion, Steering and auxiliary (Sub)	
		Operate pumping systems (bilge, ballast..)	**AEMC system**			
	Electrical, electronic and control engineering	Operate alternators, generators and control systems	**AEMC system**		Power generation and distribution system (Sub)	
	Maintenance and repair	Maintain marine engineering systems including control systems			**Maintenance planning system**	
	Controlling the operation of the ship and care for persons on board	Assigned under deck department				

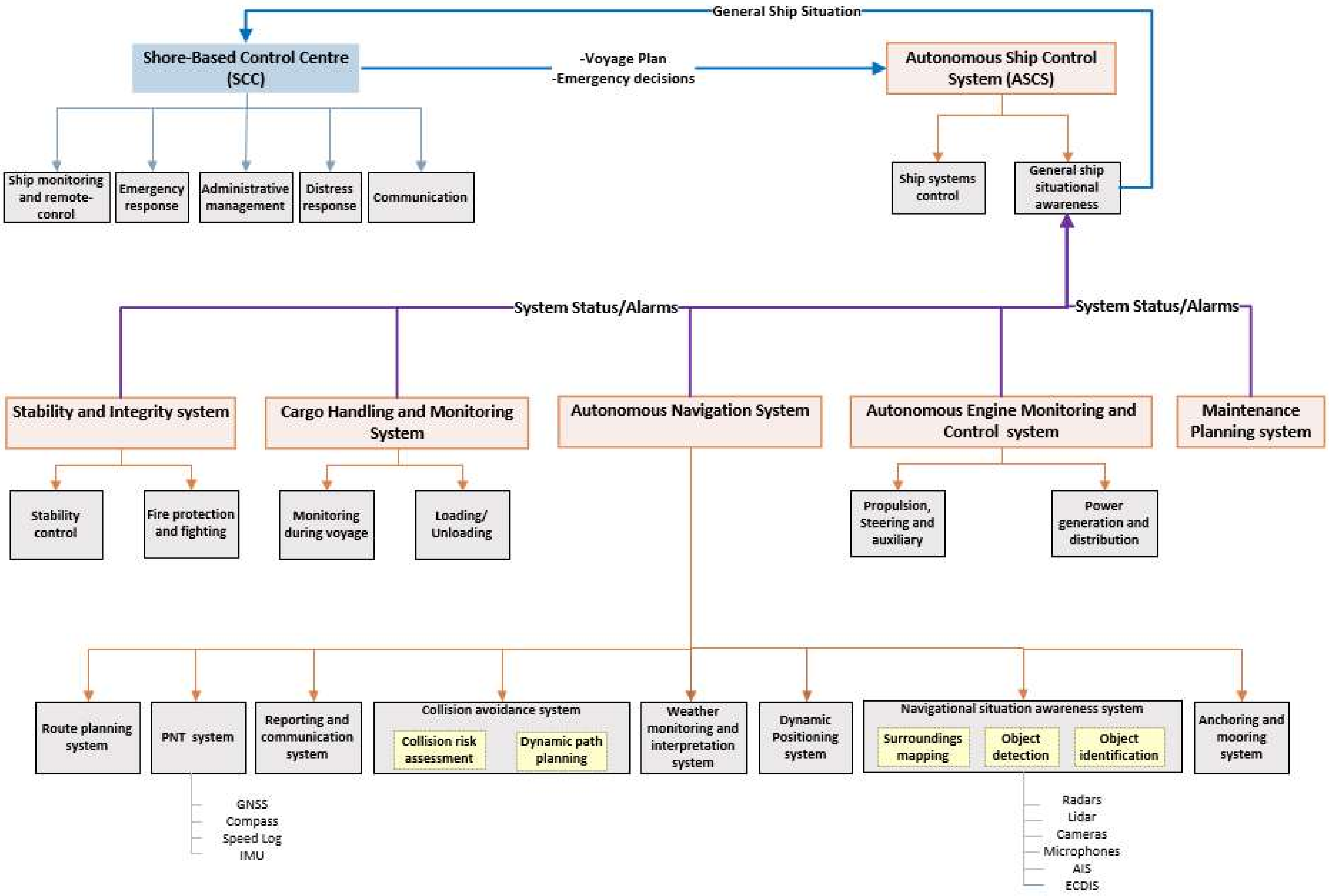

General Ship Situation
Shore-Based Control Centre (SCC)
-Voyage Plan
-Emergency decisions
Autonomous Ship Control System (ASCS)
Ship monitoring and remote-conrol
Emergency response
Administrative management
Distress response
Communication
Ship systems control
General ship situational awareness
System Status/Alarms
System Status/Alarms
Stability and Integrity system
Cargo Handling and Monitoring System
Autonomous Navigation System
Autonomous Engine Monitoring and Control system
Maintenance Planning system
Stability control
Fire protection and fighting
Monitoring during voyage
Loading/ Unloading
Propulsion, Steering and auxiliary
Power generation and distribution
Route planning system
PNT system
Reporting and communication system
Collision avoidance system
Collision risk assessment
Dynamic path planning
Weather monitoring and interpretation system
Dynamic Positioning system
Navigational situation awareness system
Surroundings mapping
Object detection
Object identification
Anchoring and mooring system
GNSS
Compass
Speed Log
IMU
Radars
Lidar
Cameras
Microphones
AIS
ECDIS

Figure 3: The proposed hierarchical systems structure of the autonomous cargo ship

5. Proposed structure

Figure 3 shows the proposed hierarchical systems structure of the autonomous cargo ship. It is obtained from the analysis described in the previous sections with technical information provided by the mentioned research projects. However, the structure is an initial suggestion of the functional boundaries of the autonomous ship systems and could be subject of further investigation in future work. More details under each function will vary by cargo type, especially for the cargo handling and monitoring system. In addition, autonomous ships other than cargo ships could have additional systems that perform other functions depending on the type of the vessel.

In this structure, each orange box represents one of the autonomous ship systems, the smaller grey boxes are their respective sub-systems. The blue box in the figure represents the SCC with its sub-systems under it with grey boxes. The name of each system refers to the function it performs as specified in the Table 1.

The structure presents the ship systems until two levels of hierarchy: systems and their sub-systems. Only the collision avoidance and the navigational awareness sub-systems have been further detailed into a third level of the hierarchy. Their respective sub-systems were added in the structure in yellow boxes and with the name referring to their functions as suggested in AAWA description of the ANS. Under the PNT and the navigational situation awareness systems, the components as suggested in MUNIN and AAWA were added in text without boxes.

The Autonomous Ship Control System (ASCS) is placed on top of the other systems in the figure only due to its control over the other systems, without any link to the principle of hierarchy in the systems theory. The ASCS is the “virtual captain” that assesses the general ship situation and controls all ship systems.

The ANS sub-systems are described in the next sections.

5.1. Route planning system

This system generates the route plan based on the information in the voyage plan. The voyage plan is delivered by the SCC. The route plan includes the way points, the speed and the heading from point to point.

5.2. Positioning, Navigation and Timing (PNT) system

This system will provide PNT information that will be distributed to other ship systems for different purposes. It will employ the technologies of the Global Navigation Satellite Systems (GNSS) receivers with satellite-based augmentation systems for better accuracy and integrity (Cueto-Felgueroso, 2018; MUNIN, 2015). Other navigation sensors such as the speed log, the Compass and the Inertial Measurement Unit (IMU) could provide the PNT information for redundancy (MUNIN, 2015).

5.3. Reporting and communication system

This system will be responsible for the automatic reporting to the shore and the Automatic Identification System (AIS). It will also conduct simple automatic communications with the ships in a collision avoidance condition. The complex communications and the distress communications will be a responsibility of the SCC.

5.4. Dynamic Positioning (DP) system

Dynamic Positioning system was suggested by Rolls Royce in AAWA project in order to steer the ship with more accuracy with the track mode having control of the ship propulsion and steering systems. The advanced DP system suggested by Rolls Royce will also have better manoeuvrability in addition to the ability of keeping a fixed position even under rough weather (AAWA, 2016). The DP system function in the proposed structure combines the “route tracking” and “manoeuvring” from Table 1.

5.5. Weather monitoring and interpretation system

This system collects the weather information from the associated shipborne sub-systems and the received weather forecast and safety warnings from the shore. It interprets this information to determine their effect on the other ship systems performance, such us visibility for the navigational situational awareness system.

5.6. Collision avoidance system

The collision avoidance system should avoid collisions in different encounter situations with conventional, remote-operated or autonomous ships. It should act according to COLREGs convention; the rule of the road in maritime traffic (AAWA, 2016). It should assess the risk of collision with the identified targets and generate a collision avoidance path with respect to COLREGs (AAWA, 2016; Perera et al., 2015; Varas et al., 2017; Lyu & Yin, 2019). It includes then two sub-systems in Figure 3 with these functions.

5.7. Navigational situation awareness system

The navigational situational awareness system merges the raw data from different sensors' readings including the traditional navigation equipment such as Radar and AIS, and the advanced sensors designed for the autonomous navigation such as infrared cameras or Lidars. As a sub-system of the ANS, it provides the situation awareness of the ship vicinity for the navigational purpose. This system will conduct the surroundings mapping to create a representation of the ship vicinity (AAWA, 2016). In addition, it should detect and identify the objects in the ship vicinity to compensate the absence of humans on board for the lookout function (AAWA, 2016). The surroundings mapping, object detection and object identification could be sub-systems of the navigational situation awareness system. The set of the components technologies under this system was proposed by AAWA project.

5.8. Anchoring and Mooring system

This system would conduct special functions of anchoring and mooring. Depending on the operational conditions, the system could be also operated by the SCC.

6. Discussion

The proposed hierarchical systems structure has included the functions of the autonomous ship under various systems that work together to steer the autonomous ship between ports.

The analysis presented the development trends in this research field as it included the ANS proposed by a more recent project than MUNIN. The route planning system in the ANS is more comprehensive than the weather routing proposed in MUNIN because it will generate a route that considers not only the weather conditions as a constraint but also other voyage plan data. Moreover, with the DP having the tracking and manoeuvring functions, it could maintain the ship position in extreme emergencies and avoid bad consequences (AAWA, 2016). This could be one scenario of the fail to safe mode, when the ship is out of control.

The analysis has also considered the experience gained in conventional ships operation by including the functions prescribed in the STCW convention. The same convention was also one important standard that helped to develop the first technical concept in MUNIN project.

The proposed structure in this paper considers the functional characteristic of each system, which means that every system and sub-system was given a function, rather than focusing on the physical boundaries of the systems. In MUNIN, the autonomous ship description was a mixture of the functions and the physical boundaries of the systems as the researchers were testing the feasibility of the concept on an existing conventional ship. The Bridge Automation for example was considered as a system while it was a mixture of components belonging to different systems. However, focusing on the physical boundaries when describing the autonomous ship systems that are under development could limit the early assessment of the safety of these systems.

The proposed structure suggests that the situational awareness module in AAWA could be called a navigational situation awareness as it ensures the awareness of the ship surroundings for the navigation. The situation awareness in its extended definition is not limited to knowing what is going around (Endsley, 2019). It should also include the awareness of the status of the ship stability, machinery, cargo and other systems that could affect the ship predefined route plan or take it into an emergency (Queensland Government, 2016). This was the function of the ship-state definition module in AAWA and of the ASCS in MUNIN. Therefore, the same function could be assigned to a sub-system (ASCS). It will receive the status and alarms data from each system and process it to provide concise situational awareness data for the ASCS decision-making. Moreover, a concise situational awareness data could be transferred to the SCC especially that the connectivity bandwidth is limited and not ready for huge amount of data (Hoyhtya et al., 2017).

On the other hand, the limited details about the technologies to be installed makes the structure missing the technical components of each system. Rolls Royce already suggested some of these components for the navigational situation awareness and many other components installed on-board currently operated ships are supposed to be part of the autonomous ships. In the future, when further details will be available about the technical components of each of the autonomous ship systems, they could be added to the proposed structure. A refined systemic hazard analysis could be then applied, which is an effective process of designing new complex systems or improving existent systems (Leveson, 2011).

The proposed structure in this study does not include the links between different systems and sub-systems as those are still not specified for all ship systems. It does not also give a strict prescription to the design of the autonomous ships. It rather represents the current development in the autonomous shipping in a structure that would be useful for systemic hazard analysis. The results of a hazard analysis based on this structure would then provide the recommendations for the design process.

7. Conclusion

In software intensive systems such as the autonomous ship systems, software could control components that contribute to the function of the system but are out of the same physical boundaries. The functional characteristic of systems as described in the systems theory was the focus of this paper. The proposed systems structure is based on available autonomous ship systems description from major research projects and known human functions from conventional ship operations.

The hierarchy was developed for conducting future systemic hazard analysis of the autonomous ship under development and contributing to its safety-based design. Deeper analysis of experienced seafarers' tasks in combination with a systemic hazard analysis technique would allow the identification of the interactions between each system components. That would also help to identify the hazards emerging with these interactions. In addition, the same approach if applied to the autonomous ship as a whole system would identify the interactions between the different ship systems and their associated hazards.

Acknowledgments

This research has been funded by RESET project. RESET has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 730888.

References

AAWA. (2016). *Remote and autonomous ships the next steps*. Retrieved from https://www.rolls-royce.com/~/media/Files/R/Rolls-Royce/documents/customers/marine/ship-intel/aawa-whitepaper-210616.pdf

Ackoff, R. L. (1971). Towards a System of Systems Concepts. *Management Science*, *17*(11), 661–671. https://doi.org/10.1287/mnsc.17.11.661

Adams, K. MacG. (2011). Systems principles: Foundation for the SoSE methodology. *International Journal of System of Systems Engineering*, *2*(2/3), 120. https://doi.org/10.1504/IJSSE.2011.040550

Altabbakh, H., AlKazimi, M. A., Murray, S., & Grantham, K. (2014). STAMP – Holistic system safety approach or just another risk model? *Journal of Loss Prevention in the Process Industries*, *32*, 109–119. https://doi.org/10.1016/j.jlp.2014.07.010

Arnold, R. D., & Wade, J. P. (2015). A Definition of Systems Thinking: A Systems Approach. *Procedia Computer Science*, *44*, 669–678. https://doi.org/10.1016/j.procs.2015.03.050

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, *253*(1), 1–13. https://doi.org/10.1016/j.ejor.2015.12.023

Basnet, S., Valdez Banda, O. A., & Hirdaris, S. (2019). The Management of Risk in Autonomous Marine Ecosystems – Preliminary Ideas. *Proceedings of the First International Workshop on Autonomous Systems Safety*. Retrieved from https://www.ntnu.edu/documents/139785/1283738018/Proceedings+of+the+1st+IWASS.pdf/dadf6629-ef88-4e48-9c2a-8576c0379da8

C. Bruhn, W., Burmeister, H.-C., T. Long, M., & A. Moræus, J. (2014). *Conducting look-out on an unmanned vessel: Introduction to the advanced sensor module for MUNIN's autonomous dry bulk carrier*. Retrieved from https://www.researchgate.net/publication/265716495_Conducting_look-out_on_an_unmanned_vessel_Introduction_to_the_advanced_sensor_module_for_MUNIN's_autonomous_dry_bulk_carrier

Cueto-Felgueroso, G. (2018). *SEASOLAS Final Report*. 63.

Endsley, M. R. (2019). Situation Awareness in Future Autonomous Vehicles: Beware of the Unexpected. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (Vol. 824, pp. 303–309). https://doi.org/10.1007/978-3-319-96071-5_32

Fleming, C. H., Spencer, M., Thomas, J., Leveson, N., & Wilkinson, C. (2013). Safety assurance in NextGen and complex transportation systems. *Safety Science*, *55*, 173–187. https://doi.org/10.1016/j.ssci.2012.12.005

Hollnagel, E. (2004). *Barriers and Accident Prevention*.

Hoyhtya, M., Huusko, J., Kiviranta, M., Solberg, K., & Rokka, J. (2017). Connectivity for autonomous ships: Architecture, use cases, and research challenges. *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, 345–350. https://doi.org/10.1109/ICTC.2017.8191000

IMO. (2011). *International Convention on Standards of Training, Certification and Watch-keeping for seafarers*.

IMO. (2018). IMO takes first steps to address autonomous ships. Retrieved August 12, 2019, from http://www.imo.org/en/MediaCentre/PressBriefings/Pages/08-MSC-99-MASS-scoping.aspx

Ishimatsu, T., Leveson, N. G., Thomas, J. P., Fleming, C. H., Katahira, M., Miyamoto, Y., … Hoshino, N. (2014). Hazard Analysis of Complex Spacecraft Using Systems-Theoretic

Process Analysis. *Journal of Spacecraft and Rockets*, *51*(2), 509–522. https://doi.org/10.2514/1.A32449

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, *42*(4), 237–270. https://doi.org/10.1016/S0925-7535(03)00047-X

Leveson, N. (2011). *Engineering a Safer World, <systems Thinking Applied to Safety*.

Lyu, H., & Yin, Y. (2019). COLREGS-Constrained Real-time Path Planning for Autonomous Ships Using Modified Artificial Potential Fields. *The Journal of Navigation*, *72*(3), 588–608. https://doi.org/10.1017/S0373463318000796

MUNIN. (2015). *MUNIN-D8-6-Final-Report-Autonomous-Bridge-CML-final*.

MUNIN. (2016). Final Report Summary—MUNIN (Maritime Unmanned Navigation through Intelligence in Networks) | Report Summary | MUNIN | FP7 | CORDIS | European Commission. Retrieved June 26, 2019, from https://cordis.europa.eu/project/rcn/104631/reporting/en

Patriarca, R., Di Gravio, G., Costantino, F., & Tronci, M. (2017). The Functional Resonance Analysis Method for a systemic risk based environmental auditing in a sinter plant: A semi-quantitative approach. *Environmental Impact Assessment Review*, *63*, 72–86. https://doi.org/10.1016/j.eiar.2016.12.002

Perera, L. P., Ferrari, V., Santos, F. P., Hinostroza, M. A., & Guedes Soares, C. (2015). Experimental Evaluations on Ship Autonomous Navigation and Collision Avoidance by Intelligent Guidance. *IEEE Journal of Oceanic Engineering*, *40*(2), 374–387. https://doi.org/10.1109/JOE.2014.2304793

Queensland Government. (2016, August 31). Situational awareness. Retrieved August 16, 2019, from https://www.msq.qld.gov.au/Safety/Situational-awareness

Rødseth, Ø. J., Kvamstad, B., Porathe, T., & Burmeister, H. (2013). Communication architecture for an unmanned merchant ship. *2013 MTS/IEEE OCEANS - Bergen*, 1–9. https://doi.org/10.1109/OCEANS-Bergen.2013.6608075

SRA. (2015). *Risk Analysis Foundations*. Retrieved from https://www.sra.org/sites/default/files/pdf/FoundationsMay7-2015-sent-x.pdf

Teikari, O. (2014). *Hazard analysis methods of digital I&C systems*. Retrieved from VTT Technical Research Centre of Finland website: http://www.vtt.fi/inf/julkaisut/muut/2014/VTT-R-03821-14.pdf

Utne, I. B., Sørensen, A. J., & Schjølberg, I. (2017). Risk Management of Autonomous Marine Systems and Operations. *Volume 3B: Structures, Safety and Reliability*, V03BT02A020. https://doi.org/10.1115/OMAE2017-61645

Valdez Banda, O. A., & Goerlandt, F. (2018). A STAMP-based approach for designing maritime safety management systems. *Safety Science*, *109*, 109–129. https://doi.org/10.1016/j.ssci.2018.05.003

Valdez Banda, O. A., Kannos, S., Goerlandt, F., van Gelder, P. H. A. J. M., Bergström, M., & Kujala, P. (2019). A systemic hazard analysis and management process for the concept design phase of an autonomous vessel. *Reliability Engineering & System Safety*, *191*, 106584. https://doi.org/10.1016/j.ress.2019.106584

Varas, J. M., Hirdaris, S. E., Smith, R., Scialla, P., Caharija, W., Bhuiyan, Z. I., … Rajabally, E. (2017). *MAXCMAS project: Autonomous COLREGs compliant ship navigation*.

Wróbel, K., & Montewka, J. (2019). Comments to the article by Ramos et al. ‘Collision avoidance on maritime autonomous surface ships: Operators’ tasks and human failure events’ (Safety Science Vol. 116, July 2019, pp. 33–44). *Safety Science*. https://doi.org/10.1016/j.ssci.2019.03.024

Wróbel, K., Montewka, J., & Kujala, P. (2017). Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. *Reliability Engineering & System Safety*, *165*, 155–169. https://doi.org/10.1016/j.ress.2017.03.029

7th European STAMP Workshop & Conference
18 - 20 September 2019, Helsinki

Development of functional safety requirements for DP-driven servicing of wind turbines

Romanas Puisa[1,*], Victor Bolbot[1] and Ivar Ihle[2]
[1] Maritime Safety Research Centre, University of Strathclyde, UK
[2] Kongsberg Maritime, Norway

ABSTRACT

The adage "prevention is better than cure" is at the heart of safety principles. However, effective accident prevention is challenging in complex, highly automated systems such as modern DP-driven vessels, which are supposed to safely transfer technicians in often unfavourable environmental conditions. FMEA analysis, which is required for DP-driven vessels, is helpful to build-in a necessary level of redundancy and thereby mitigate consequences of failures, but not particularly helpful to inform preventive measures, not least against functional glitches in controlling software. In this paper we develop a set of functional safety requirements which are aimed at prevention of causal factors behind drift-off, drive-off and other hazardous scenarios. For this purpose, we use a systemic hazard analysis by STPA, which delivers both failure and interaction-based (reliable-but-unsafe) scenarios. The functional requirements cover both design and operational (human element related) requirements, which are then ranked based on our proposed heuristic. The ranking is not predicated on statistics or expert option but instead it is proportional to the number of hazardous scenarios a requirement protects against, hence indicating the relative importance of the requirement. The paper also summarises the suggested areas of safety improvement for DP-driven vessels.

Keywords: windfarm; wind turbine; dynamic positioning; service offshore vessel; technician transfer

1. INTRODUCTION

1.1 SERVICE OFFSHORE VESSELS

Offshore wind-framing is becoming a major source of renewable energy in many countries. As wind farms are moving further offshore, significant innovations in the infrastructure and services are required to maintain the judicious trend. One of such innovations is the specialised service vessels, or service offshore vessels (SOVs), which are offering new logistical concepts for servicing windfarms further offshore. They enable an extended stay of technicians (typically for two weeks) in the vicinity of a windfarm, thereby replacing the logistical concept of technician transfer from shore. The latter becomes unreasonable due to prolonged sailing times and increased risk of seasickness. SOVs, which are typically around 90 meters in length, can also endure more severe environmental conditions and offer a wide array of services. They are smart ships (highly automated), hosting dozens of technicians, heavy equipment and means of its handling. SOVs are also complex systems with many components (some subsystems are partly autonomous) and layers of communication between them.

---

[*] Corresponding author: +44 141 548 32 45 and r.puisa@strath.ac.uk

There are various ways of how the SOV can be utilised, and depends on specific circumstances (current and future) of a windfarm. In some cases, the SOV can be the only vessel at a windfarm to transfer technicians and equipment. In others, it can be part of a bigger fleet of vessels of various sizes and functions; a SOV would normally interact with all players in the fleet. Such a fleet, for instance, can comprise a SOV, daughter crafts, and a floatel (floating hotel). The latter is well suited for technicians and crew to be resting on undisturbed, when the other vessels are serving turbines 24/7. Daughter crafts (DCs) are medium size boats (under 20 meters) which are carried by the SOV and used to transport lighter equipment to turbines in moderate environmental conditions (< 1.8m significant wave height). DCs are loaded with technicians and launched from a SOV deck by some davit system (typically 3-5 times per day) and then recover (lift up) DCs from the water. SOVs would also have a sophisticated system for transferring technicians and equipment to and from a turbine. It is normally a motion-compensated (3 or 6 DoF) gangway which allows for the safest (based on experience so far) and time-efficient (within 5 minutes) transfer.

Regardless a logistical concept selected for a given windfarm, there are a number of functional requirements that a SOV has to fulfil. One of them is station keeping, i.e. the ability to maintain position and heading within their tolerable ranges and for an extended period of time under all operational conditions. Another is the ability to strictly follow a predefined trajectory along waypoints. These two functions are needed for both productivity (the number of turbines serviced per unit of time) and safety (prevention of injury and death among crew and technicians). The key system that provides these functions is the *dynamic positioning system* (DP system). The DP system is the object of this paper.

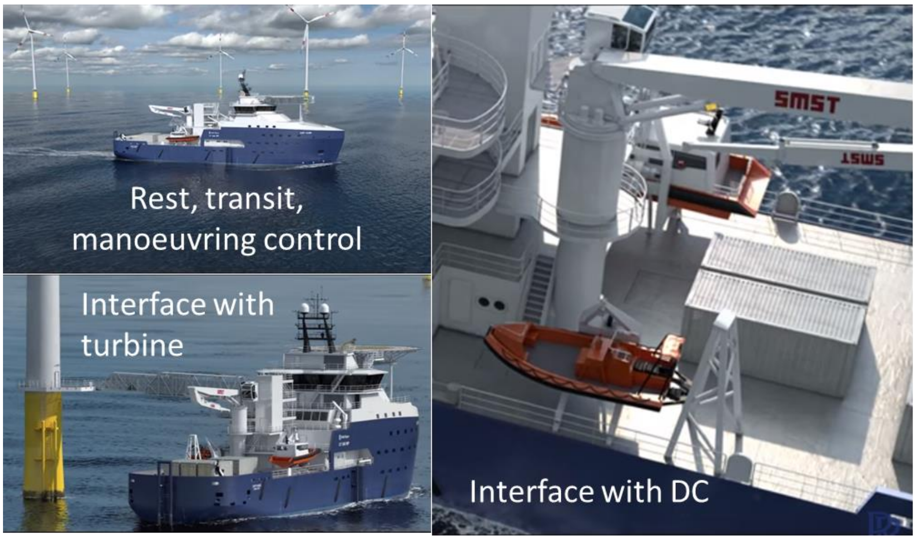


Figure 1: Operation modes when DP system is used (courtesy of Kongsberg Maritime / fmr Rolls-Royce Marine)

The DP system is, hence, involved in multiple operational modes of a SOV (cf. Figure 1). That is, when the vessel is transiting from shore to a windfarm, resting (night time) with people onboard, manoeuvring between turbines, and interfacing with turbines or daughter crafts. These modes of operation are safety critical and there are different safety hazards to watch for. For instance, during a transit or manoeuvring, the vessel might collide with turbines or other vessels, e.g. when the vessel deviates from a correct trajectory or inadequately performs collision avoidance. This can happen even in the area of a windfarm where fishing and other vessels are allowed to enter, as the case in the UK and other nation states. The loss of position or heading due to drift-off or drive-off scenarios are primary hazards during the resting and interfacing modes. Drift-off is a situation of the vessel drifting away after a loss of thruster power, whereas drive-off happens when the vessel is being pushed away by excessive thruster force.

1.2 SAFETY ASSURANCE AND ITS DEFICIENCIES

These safety hazards are normally pre-empted by ensuring a necessary level of reliability of station (position and heading) and trajectory keeping functions. Reliability of critical sub-systems and components is achieved through their redundancy. The vessel can operate at a different level of reliability (aka DP-equipment class (IMO, 1994)), depending on the safety criticality of a current mode of operation. For instance, DP-equipment class 1 (DP1) does not require redundancy and would normally be used when the vessel is resting, transiting and manoeuvring within so-called safe zones. In turn, DP2 and DP3 would be used in other operational modes where station keeping is key, e.g. technician transfer to or from a turbine. Therefore, DP2 and DP3 require redundancy against single failures of active and statics components such as generators, thrusters, valves, cables etc. Such single failures also include inadvertent acts by the people onboard the vessel. Currently, the main design and verification method of sufficient redundancy is the failure mode and effect analysis (FMEA) (DNVGL, 2015; IMCA, 2015). Other operational hazards, including those occurring when the vessel is in DP mode, are essentially left to be "managed by vessel operators as part of their safety management system." (IMCA, 2015).

However, although this approach to achieving safety is necessary, it is insufficient in several aspects. Firstly, ensuring reliability of both technology and people does not guarantee safety in complex systems, and can even be iatrogenic (Besnard & Hollnagel, 2014; N. G. Leveson, 2011). Complex systems feature *complex* interactions between system components, i.e. "the interactions in an unexpected sequence" (Perrow, 1984, p. 78), and accident can occur because of uncontrolled interactions of otherwise healthy components (Tiusanen, 2017, p. 464). Example interactions occur when one component is using another component when it should not or how it should not, i.e. typical cases of mode confusion. As these interactions within the entire system, safety is a system but not a component property. A related issue is that FMEA is used in a bottom-up manner, i.e. it attempts to identify the effect of a component failure on system safety. This is contrary to the notion that safety is a system property. Consequently, FMEA becomes also insufficient, for it is fundamentally biased towards accident scenarios caused by component failures and discounts those caused by dysfunctional interactions, i.e. system design errors. Secondly, one cannot foresee all interactions (and effects thereof) in complex systems, and hence a safety analyst should focus on improving control of component interactions at the functional level, as opposed to physical level where FMEA would normally operate at. Thirdly, FMEA would also misinterpret the contribution of people and software to accident scenarios (Victor Bolbot et al., 2018), for neither people nor software can credibly be said to fail rather than merely following wrong instructions (Dekker, 2014; N. G. Leveson, 1995).

1.3 CONTRIBUTION

Given these deficiencies of the current approach to safety of DP-driven vessels, we applied an alternative one. It is based on the method of systems theoretic process analysis (STPA) (N. Leveson, 2011; N. Leveson & Thomas, 2018). The method allowed addressing the highlighted deficiencies of failure-based analysis by FMEA and end up with functional safety requirements, which can be used by both system designers (e.g., software developers and integrators) and operators as part of their safety management systems. STPA is a hazard analysis method and it, hence, targets the initial phase of risk assessment, namely the hazard identification and analysis (ISO 31000, IEC/ISO 31010). The paper explains how we performed the STPA analysis of the DP system within various modes of SOV operation (cf. Figure 1), specifically focusing on hazards, analysis process, development of functional requirements, and result communication.

The latter conventionally requires to quantitatively rank individual scenarios identified through hazard analysis, essentially following the bottom-up approach. This was found especially challenging, given that the information about individual scenarios is scant (unreliable) or absent. We, hence, developed a heuristic to bypass this difficulty: instead of scenarios (pathways to system hazards), functional requirements against these scenarios were ranked. The used approach is congruent with the systems thinking that underpins STPA.

STPA has been applied to DP-driven vessels before, e.g. (Abrecht & Leveson, 2016; Rokseth et al., 2017). However, the analysis presented in this paper addresses different operational context and modes of operation (e.g., SOV interfacing with a turbine), and covers scenarios excusive to SOV servicing of windfarms. The paper does not explain the STPA method and expects the reader to be conversant with it. The unfamiliar reader is referred to the STPA handbook (N. Leveson & Thomas, 2018).

The paper is organised in two parts. The first part explains the assumptions behind the hazard analysis by STPA. Essentially, it explains what has been done, how and why. The second part summaries the analysis results in terms of high-level requirements, and concludes the paper.

2 ANALYSIS ASSUMPTIONS

This section covers essential assumptions behind the hazard analysis process by STPA. These assumptions concern about the system analysed, its objectives and hazards, generation of hazardous scenarios and corresponding functional requirements for their prevention and mitigation. The adopted approached for ranking and validation of the requirements is also discussed in this section.

2.1 SYSTEM AND ITS HAZARDS

As explained in the introduction an SOV is a highly-automated and multifunctional vessel. The DP-system is used in various, fairly mutually exclusive, modes of SOV operation and interaction with other objects in a windfarm (cf. Figure 1). The overall system of such interactions is shown in Figure 2. The analysis covered the five interactions whose safety is affected by the DP system. These interactions are of physical contact (e.g., SOV and turbine), communication via radio (e.g., SOV and shore, turbine and shore), and sensory (distance, visual, and audio) by installed sensors and people. Other interactions at the system level (i.e. the links between the DC and turbine or other ships) were not analysed.

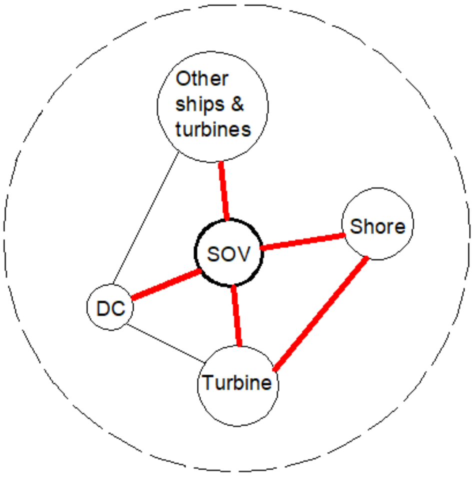


Figure 2: System components and system boundary

Figure 3 shows a simplified version of hierarchical control diagram with the DP control system involved. The human operator (HO) acts as the top controller and there are essentially four modes of interaction with the DP system:

1. DP system is in auto mode. DP autonomously achieves position, heading, or trajectory setpoints, whereas the role of HO is only supervisory with the ability to intervene when required. DP can also automatically switch thrusters to manual control by levers if failure or other anomalies are detected.

2. DP system in joystick mode. HO can control certain vessel axes (sway, surge or yaw) with DP controlling others. DP can also switch thrusters to manual control as described above, in turn HO can ask DP to take over control of manually controlled axes.
3. DP system controls some axes only. HO uses manual levers to control specific thrusters.
4. DP system is not controlling thrusters and it is either in standby or disabled mode. HO controls thrusters by manual levers.

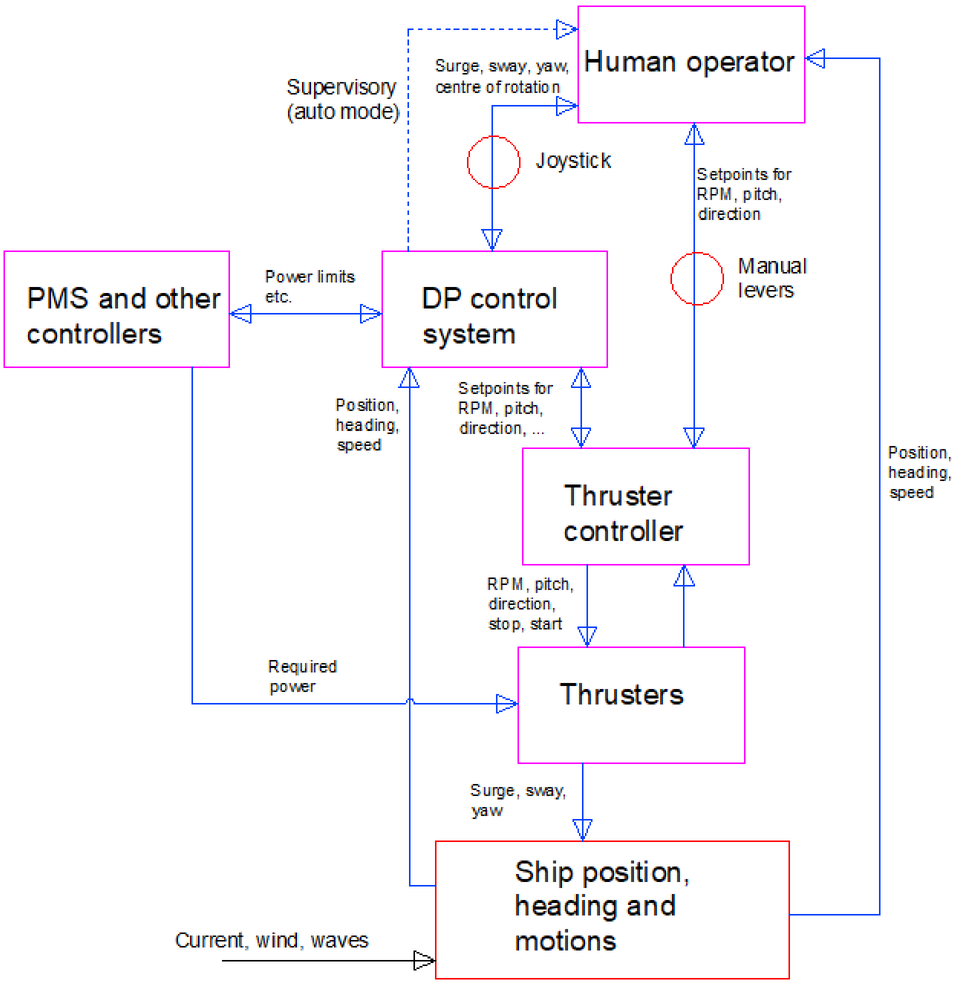


Figure 3: High-level representation of DP control and other systems (only a part of control and feedback information is displayed; some control and feedback channels are joined for simplicity)

During the transit mode, the SOV can either be in auto pilot (i.e., DP controls thrusters by following waypoints) or manual (joystick or levers). During manoeuvring between turbines (incl. turbine approach and departure), all axes of the SOV would normally be controlled by joystick. However, an autonomous manoeuvring would also be possible on novel vessels, when the SOV would autonomously approach a turbine, unload/load technicians and equipment via a gangway, and depart. In this case, the DP system will need to have this function. During interfacing with a turbine or DC, the SOV is supposed to keep station (position and heading) and this is usually done by the DP system being in auto mode (i.e., controlling all axes).

The control diagram in Figure 3 also shows other controllers such as the power management system (PMS). The interactions between these systems were included in the presented analysis, however PMS hierarchy and other systems were analysed in a separate study also presented in this conference (V. Bolbot et al., 2019).

Once the system has been defined, the next step is to formulate accidents (undesirable losses) and system-level hazards (how these losses can occur). The used rule of thumb, when formulating accidents and system hazards, was that accidents would correspond to undesirable deviations from or disturbances to the prime system objective (this formulation agrees with the definition of risk in ISO 31000), whereas hazards would essentially correspond to violated constraints which are necessary to achieve the objective. For instance, the prime system objective is to safely transfer technicians and equipment in minimal time (or minimal fuel consumption rate) and across a range of prescribed environmental conditions. Requirements and constraints to achieve this objective correspond to availability of adequate capacity of engineering systems (e.g., DP, davit) and adequate interactions between technology and people. Specific violations (or disregard) of such requirements and constraints, would allow formulating the hazards such as drifting off or driven off the position or hearing. Thus, in our case the accidents in question are: (A1) Injuries or loss of life, (A2) damage or loss of ship or other assets (daughter craft, gangway, davit system, or turbine). Table 1 list system hazards considered in various modes of operation. Note, only hazards related to the DP system and the interaction between DP and HO are shown, whereas other hazards (e.g., the gangway is retracted while in use by technicians) were also considered but are outside the scope of this paper. Some of the listed hazards were informed by current safe rules and recommendations such as IMO COLREGS (safe navigation), IMCA MSF (safe operation of DP, (IMCA, 2015)), etc.

Table 1 System hazards

Mode of operation	System hazards
Transit	H1: Sailing and stopping (crash stop) within a distance appropriate (minimal safe distance) to the prevailing circumstances and conditions (other ship, turbine etc.) is not achieved. H2: Ship course does not change promptly to avoid collision (astern, forward, sway, yaw). H3: Large and observable alteration of course are not achieved (as opposed to small alterations).
Manoeuvring between turbines (incl. turbine approach and departure)	H1 H4: Required course cannot be maintained for predefined time (on autopilot / DP / manual).
Rest	H5: Position and/or heading is not maintained (drive-off, drift-off) within the predefined ranges before an operation is completed. H6: Station keeping capability does not match the operational requirements of the vessel.
Interface with turbine	H5, H6
Interface with daughter craft	H5, H6

The control diagram in Figure 3 was analysed by considering four separate loops: HO-joystick-DP, HO-levers-Thruster Controller, DP-Thruster Controller, Thruster Controller-Thrusters. The system hazards were decomposed into loop-related sub-hazards to facilitate the local analysis. For instance, the loop Thruster Controller-Thrusters had the following sub-hazards:

- H5.1: Setpoints are not achieved in required time.
- H5.2: Setpoints are not maintained within alarm limit.
- H5.3: Communication between thruster and remote controller is not maintained at required frequency.
- H5.3: Loading of el. motors and/or diesel engines exceeds the limits.

Table 2 summarises control actions per control loop. The list of analysed control actions is helpful to grasp the scope and detail of the analysis.

Table 2 Summary of control actions per control loop

Control loop	Control actions
HO-joystick-DP	• Update setpoint (sway, surge, yaw, or all) • Change joystick device gain for manual heading/position/rotation and thrust bias (low, medium, high) • Change axis control mode (auto, joystick, no control/levers) • Change centre of rotation • Change DP control mode (relaxed, normal) • Change vessel control mode (manual/auto position, auto/manual sway, auto/manual surge, manual/auto heading, trajectory) • Change vessel draught in operation monitor panel (for auto heading) • Change alarm/warning limits (4/5m for default warning/alarm) • Wait until DP settles (20 min) • Release to Manual • Change operational objective/task • Change IMO DP class
HO-levers-Thruster Controller	• Start thruster (make thruster ready to use; lever in command) • Update setpoint (RPM, pitch, direction) • Enable/disable thruster • Stop/shutdown thruster • Control transfer (transfer command between bridge and engine control room) • Command transfer (take command from other controllers of thruster. Make the lever in command)
DP-Thruster Controller	• Update setpoint for individual thrusters or thruster group (RPM, pitch, direction, moment, timing/acceleration) • Enable thrusters • Disable thrusters • Take control of axis • Release control of axis
Thruster Controller-Thrusters	• Acknowledge communication signals from remote control system • Achieve setpoint (RPM, pitch, direction) • Maintain load control (azimuth thrust controllers)

2.2 HAZARDOUS SCENARIOS

We used the STPA process described in (N. Leveson & Thomas, 2018) to come up with hazardous scenarios, i.e. combinations of unsafe control actions (UCAs) and their causal factors (CFs). The identification of UCAs and CFs was done manually, and with the guidance of the conventional modes and guidewords for UCAs (e.g., control action is not provided, wrong provided, provided too late, etc.) and CFs (e.g., inconsistent process model, out-of-range disturbances etc.); see (N. Leveson & Thomas, 2018).

Formulation of potential CFs is generally more challenging than of UCAs. It is particularly strenuous when it comes to human controllers, as opposed to automated counterparts. CFs for the latter were addressed by answering the following guiding questions:

- What process model (PM) would cause a given UCA?
- How such a PM would be created?
- How PM should be interpreted to cause the UCA?
- How the control action should be executed to cause UCA?

For human controllers (e.g., human operator controlling the vessel position by manual levers), similar questions could be asked (although replacing PM by the mental model). Additionally, we used a list of further guidewords grouped into four phases of decision/action making on the part of human: observing/receiving information, interpreting and updating the mental model, deciding on specific action, and executing action. Some of the generic causal scenarios of how each group can be undermined are shown in Table 3; comments on these scenarios are found in (Bainbridge, 1983; Hollnagel, 2017; Lee, 2008; N. Leveson, 2011; N. G. Leveson, 1995; Sarter et al., 1997). These generic scenarios can be regarded as templates for specific causal factors which reflect the context at hand.

Table 3 Sample guidewords for formulating causal factors for human controllers

Function	**How this function can be undermined**
1. Observing / receiving	• Clarity of information (display design, visual destructions etc.): information is unnoticed, noticed too late or misunderstood • Low alertness, monotonicity of process: information is unnoticed or noticed too late • Graceful failure of automation: information is unnoticed or noticed too late • Supra commands (missing, wrong, untimely): no relevant and timely information from top controllers • Operator is unskilled and over loaded: automation requires more skilful and less loaded operator for effective reaction in emergencies • Controlled software/process does not provide adequate feedback on operator errors, who hence does not notice them or notice too later (in life such feedback often instant and clear) • Tunnel vision (extreme fear or distress, most often in the context of a panic attack, sleep deprivation): information is incomplete or wrong • Control panel displays change unexpectedly and to a different, less familiar one / operator is used to some display, but it changes to different one in emergency: information is unnoticed, ignored or misinterpreted • Uncertain default settings which do not change with operational modes (difficult to know when the settings are hazardous): crucial information is ignored, misinterpreted
2. Interpreting and updating mental model	• Mode confusion when modes change automatically/autonomously, seamlessly, without warning: crucial information is unnoticed, misinterpreted • Operator does not know what task the computer is dealing with and how (unclear allocation of responsibilities): information is misinterpreted, ignored • Nondeterministic automation, irregular, unpredictable behaviour: information is misinterpreted or ignored (e.g. assuming a fault or outlier) • Complacency, overreliance on automation (when automation makes no sense): information is misinterpreted or ignored • Training, experience: information is unnoticed, misinterpreted or ignored (e.g. unfamiliar factors are ignored)

Function	How this function can be undermined
	• Working storage, i.e. limited (only local) information is available just after take-over (i.e. after taking over the operator has limited info about the system state): information is misinterpreted
3. Deciding on action	• Cost-benefit trade-off (e.g., wrongly thinks it is not beneficial to do, or beneficial to do): necessary action may not be taken or delayed • Safety criticality (e.g., operator thinks it is not safety critical): necessary action is not taken or delayed • Confused accountability, responsibility with other controllers: necessary action is not taken or delayed, wrong action is taken • Unrepaired/partly repaired fault (by some other controller) is unexpectedly returned to operator (e.g., for manual control): relevant action is not found in time, action is delayed
4. Executing action	• Procrastination: execution is postponed (e.g., waiting on favourable weather) • Due to irresponsiveness etc., operator assumes a failure in automation: action is delayed, action is inadequate

2.3 FUNCTIONAL REQUIREMENTS

A function is a useful capability provided by one or more components of a system. Functional requirements describe what the system *must do* (or, formally, ‘shall do’), rather than how it must do it (Young, 2004). The latter is addressed by non-functional requirements. Functional safety requirements (incl. safety constraints at the functional level) define functions for safety barriers (or defences) to be put in place against specific hazardous scenarios. Each requirement should have a rationale, type, priority and other information to facilitate decision making by designers or operators (see for instance ISO/IEC/IEEE 29148:2018). In our work, the rationale corresponded to hazardous scenarios—combinations of UCAs, CFs and hazards—and other contextual information such as corresponding control actions, controlled processes etc.

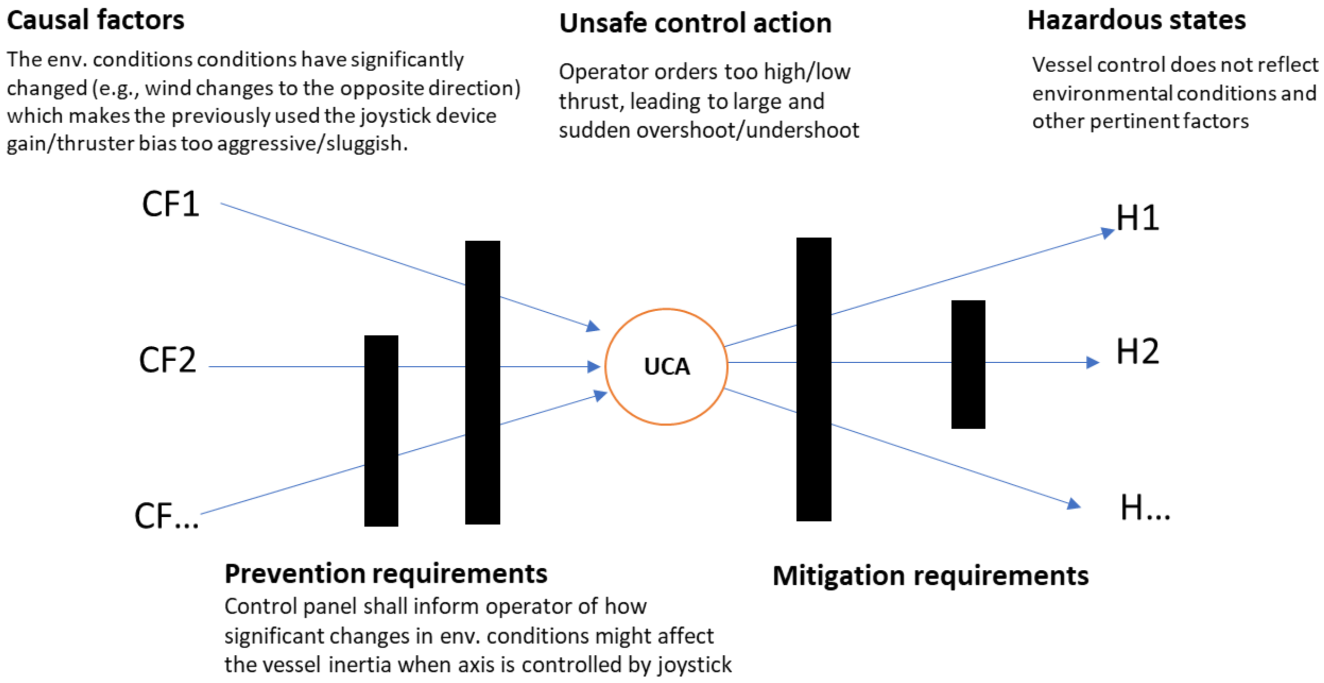

Causal factors
The env. conditions conditions have significantly changed (e.g., wind changes to the opposite direction) which makes the previously used the joystick device gain/thruster bias too aggressive/sluggish.
Unsafe control action
Operator orders too high/low thrust, leading to large and sudden overshoot/undershoot
Hazardous states
Vessel control does not reflect environmental conditions and other pertinent factors
CF1
CF2
CF…
UCA
H1
H2
H…
Prevention requirements
Control panel shall inform operator of how significant changes in env. conditions might affect the vessel inertia when axis is controlled by joystick
Mitigation requirements

Figure 4: Prevention and mitigation functional requirements (examples are provided in small print)

The derived requirements were classified into UCA prevention and mitigation requirements as illustrated in Figure 4. Prevention requirements would directly aim at causal factors, thus preventing UCAs in the first place. Mitigation requirements would react to the realisation of the UCAs, so they do not lead to hazards. Clearly, the adopted classification is subjective and relative to what we put in the middle of the bowtie. For instance, if a hazard (some hazardous system state) is in the centre, then both requirements become preventive. Both set of requirements were further classified into design and operational.

As the hazard analysis covered four control loops (cf. Section 4), requirements were primarily aimed at design and operation of controllers. Some controllers (e.g., human operator) were involved in several loops and hence contexts. That allowed to derive additional requirements for such controllers. As the number of requirements was significant, were ranked according to a heuristic described in the next section.

2.4 REQUIREMENT RANKING

A hazard analysis by STPA would normally end up with many hazardous scenarios—in our case hundreds of them—with similar number of requirements (typically smaller, for some requirements cover multiple scenarios). The myriad of requirements is obviously unconducive to the communication of hazard analysis results. Therefore, some quantitative ranking of requirements is usually adopted to alleviate this problem. Ranks would normally reflect risk-related information attached to corresponding scenarios, e.g., scenario likelihood, consequences or both.

However, the likelihood information was missing in our case. The uncertainty with scenario likelihood (or probability) is common, especially for non-standard and new technology. We were also reluctant about eliciting subjective estimates from domain experts, given how biased and unreliable outcomes could have been, e.g. (Skjong & Wentworth, 2001). The situation with scenario consequences was much simpler, for the identified scenarios led to predefined hazards and corresponding accidents.

There are, however, a number of conceptual issues with quantification of hazardous scenarios. Firstly, there is no evidence that quantification per sei improves safety (e.g., by directing resources to high risk scenarios) (Rae et al., 2012). Inaccurate estimates of associated likelihoods can be precarious, equally as navigating by a map of a wrong city. It is hence better to have no guidance at all, than the wrong one. Secondly, safety is an emergent system property, i.e. the system is not the sum of its components (Rasmussen, 1997). Hence, the assumption is that—given the emergent property of safety—there is no need to quantify individual scenarios leading to system hazards, in fact it would be incongruent with systems thinking. Quantification should only be done to system hazards—based on experience (typically supported by statistics) or expert opinion—but not to component-level scenarios.

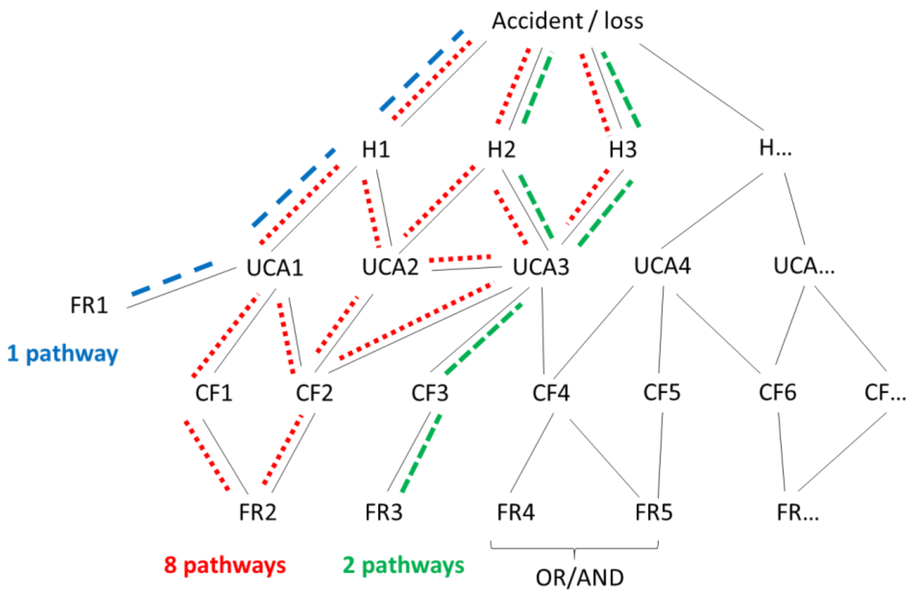


Figure 5: Accident pathways addressed by safety requirements

With the above in mind, we adopted a heuristic to rank the requirements, as opposed to hazardous scenarios directly. We took advantage of the available traceability between requirements and accidents, additionally factoring in the information on the type of requirements. Figure 5 shows a resultant tree of the hazard analysis by STPA. Functional requirements (FRs) address specific causal factors (CFs) or directly unsafe control actions (UCAs). In the latter case, the requirements would be of mitigation type (e.g., FR1). If a requirement is implemented, it blocks specific pathways to accidents in question. As shown in Figure 5, FR1 would blocks just one pathway, whereas FR2 and FR3 block 8 and 2 pathways respectively. Clearly, the importance of a requirement is proportional to the number of pathways it blocks. Note, there could be also a path from one UCA to another (e.g., UCA2 to UCA3) in the scenario tree. This path reflects the control hierarchy, i.e. UCA2 belong to a high lever controller which controls (or affects in some other way) a controller that issues UCA3. This result scenario tree is comprehensive.

In addition to the number of pathways to accident, which reflects the impact level of a requirement, the requirement type was factored in into the requirement rank. In this case, the requirement type corresponded to whether a requirement aims to prevent or mitigate a UCA. We followed the general principles compactly reflected in the adages: “prevention is better than cure” and “an ounce of prevention is worth a pound of cure”. In other words, prevention of some unfavourable events such as UCAs is more effective than their mitigation; recall the hierarchy of control by HSE (Books, 1997) and risk control by NASA (Bahr, 2014, p 29). Other figures such as the difficulty to implement a requirement (as proxy for cost) can also be added, but they are not discussed in this paper.

The requirement rank was then calculated as follows:

$$\text{Rank} = \text{Impact} \times \text{Effectivness} \qquad (1)$$

Where the impact equals to the number of pathways counted on the result tree (cf. Figure 5), whereas the effectiveness equals 1 for mitigation and 2 for prevention requirements. There is, however, one caveat to the ranking of this kind. Requirements that receive low ranks can, in principle, be equally safety critical as those of high rank. Hence, a low rank should not be the basis for discarding the requirement, but rather as an indicator that the requirement is lower in the review priority list.

Note that some requirements can be complementary (AND) or redundant (OR), as indicated in Figure 5. This information was not factored in the ranking, and is meant to be used during the later stages when requirements are fulfilled by specific safety barriers, i.e. design, operational or organisational measures.

2.5 REQUIREMENT VALIDATION

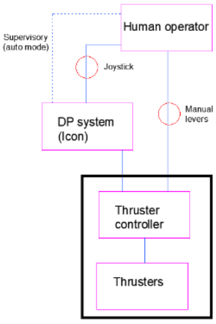
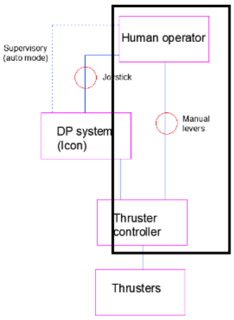
Requirements validation was performed by designers (of DP and other control systems) and experts working in design approvals. The experts were asked to review the requirements (starting with high rank ones) and corresponding scenarios, and comment on their validity, i.e. if scenarios were possible (can happen) and requirements were realistic and sound. Consequently, only valid scenarios and requirements were retained. An analogous conservative approach for scenarios filtering is advocated by Leveson (N. Leveson, 2015).

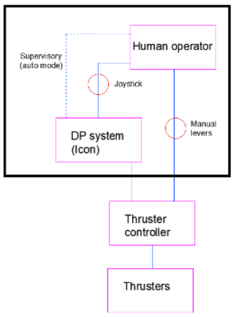
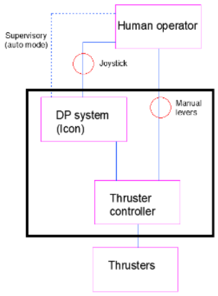
3 SUMMARY OF RESULTS

Table 4 contains sample requirements which scored highest ranks. Each requirement blocks dozens of hazardous scenarios behind vessel drift-off, drive-off and other situations. The requirements are predominantly preventive (i.e., target causal factors of UCAs) in the analysed control loops. Same requirements were derived in two control loops: FR3-5 are same as FR7-9, also FR6 is same to FR13. Consequently, these requirements have higher ranks than the others.

In summary, the functional requirements target inadequate feedback to the operator about system malfunction (and early precursors thereof) and healthy states, both of which are hazardous, as well as emergency states. In the latter case, the requirements imply the need for decision support in emergency. Some requirements such as FR19 echo the current requirements for the DP system.

Table 4 Sample requirements of high priority

Control loop (simplified versions of Figure 3)	**Design requirements**	**Operational requirements**
 Supervisory (auto mode) Human operator Joystick Manual levers DP system (Icon) Thruster controller Thrusters	1. Thrust control system shall be able to deal with external obstructions of thrusters (e.g., fishing nets, plastic waste)	2. Precautions shall be in place against manual setting of wrong load limits for el. motor and engines
 Supervisory (auto mode) Human operator Joystick Manual levers DP system (Icon) Thruster controller Thrusters	3. Indication shall be provided of malfunction criticality of thrusters (not just failed/not failed) 4. Warning of emergency situation shall be provided to operator 5. Assessment with and indication of env. effects on vessel's manoeuvrability shall be provided to operator	6. Operator shall have adequate conversancy with emergency procedures and recovery actions

Control loop (simplified versions of Figure 3)	**Design requirements**	**Operational requirements**
 Supervisory (auto mode) Human operator Joystick DP system (Icon) Manual levers Thruster controller Thrusters	7. Indication shall be provided of malfunction criticality of thrusters (not just failed/not failed) 8. Warning of emergency situation shall be provided to operator 9. Assessment with and indication of env. effects on vessel's manoeuvrability shall be provided to operator 10. Operator shall be advised with recover actions in emergency 11. Accurate visuals (using cameras etc.) of the relative vessel position/heading with respect to turbine, DC etc. shall be provided to operator	12. Timely and unambiguous communication of operational objectives to operator shall be provided 13. Operator shall have adequate conversancy with emergency procedures and recovery actions
 Supervisory (auto mode) Human operator Joystick DP system (Icon) Manual levers Thruster controller Thrusters	14. DP system shall get immediate awareness of all failure modes of thrusters/thruster controller 15. DP system shall check the entered (by operator) position/ heading/trajectory alarm limits against safety etc. criteria (i.e., sanity check) 16. DP system shall warn operator about inadequate alarm limits 17. DP system shall consider delays and irregularities in thruster signals 18. DP system shall notify operator about communication delays with thruster 19. DP system shall perform continuous assessment of the effect of environmental conditions on DP operability	20. Operator shall check the entered position /heading/trajectory alarm limits against safety etc. criteria (i.e., sanity check)

4 CONCLUSIONS

The paper has summarised a hazard analysis of a DP-driven vessel servicing windfarms which are located far offshore. The objective of the analysis was to come up with functional design and operational requirements to be used as input to a vessel design process, as well as to the development of a safety management system (SMS). The requirements were meant to be at the functional level (non-prescriptive), so designers could use them at early design stages and decide on specific safety measures that fulfil them. To this end, the hazard analysis was performed by the method of systems theoretic process analysis (STPA), which we found pertinent to achieve this objective.

The hazard analysis has focused on the DP system as it operates in various operational modes when vessel drift-off, drive-off and other hazards can happen. Hundreds of scenarios that can lead to such system hazards have been identified and used to derive functional safety requirements. The requirements were ranked by the proposed heuristic which takes advantage of the scenario tree and other aspects. The scenario tree allows to count the number of hazardous scenarios (component-level pathways to system hazards) a requirement protects against, hence indicating the relative importance of the requirement. In other words, the ranking is not predicated on scenario risk contribution, likelihood or other scenario-level information. And it is not because creditable likelihood information on hazardous scenarios in is absent in complex systems, but that quantifying individual scenarios is incongruent with the systems thinking. Hence, the proposed ranking approach matches the systemic spirit of STPA.

The paper has then summarised and discussed design and operational requirements which received high ranks. Thus, adequate feedback (timely, accurate and complete) to the bridge operator was found to be indispensable to maintain safety during technician and equipment transfers by the SOV. And improvements should be firstly directed to providing adequate:

- Feedback to the bridge operator about system malfunctions and early precursors thereof.
- Feedback on DP settings that can become hazardous in certain modes.
- Feedback when the vessel enters emergency states.
- Feedback on current and unfolding environmental conditions, and their effect on the DP and vessel performance.
- Decision support in emergency.

There are a few caveats to the study. The paper has not discussed how the requirements can be implemented or achieved, given these are only functional requirements that define functions for safety barriers but not barriers themselves. Consequently, cost effectiveness analysis of corresponding safety barriers could not be considered. The paper has not provided a detailed comparison of the derived requirements against the current requirements for the DP systems, although some high priority requirements (cf. FR19 in Table 4) echo the existing safety rules; a detailed gap analysis will be the object of a follow-up study.

ACKNOWLEDGEMENTS

The work described in this paper was produced in research project NEXUS[†]. The project has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 774519. The authors are thankful to their colleagues and project partners who directly and indirectly contributed to the presented work.

REFERENCES

Abrecht, B., & Leveson, N. (2016). Systems theoretic process analysis (STPA) of an offshore supply vessel dynamic positioning system. *Massachusetts Institute of Technology, Cambridge, MA*.

Bahr, N. J. (2014). *System safety engineering and risk assessment: a practical approach*: CRC Press.

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129-135): Elsevier.

Besnard, D., & Hollnagel, E. (2014). I want to believe: some myths about the management of industrial safety. *Cognition, Technology & Work, 16*(1), 13-23. doi:10.1007/s10111-012-0237-4

Bolbot, V., Puisa, R., Theotokatos, G., Boulougouris, E., & Vassalos, D. (2019). *A comparative safety assessment for Alternate Current, Direct Current and Direct Current with hybrid supply power systems in windfarm Service Operation Vessel using System-Theoretic Process Analysis*. Paper presented at the The 7th edition of the European STAMP Workshop and Conference (ESWC), Helsinki.

Bolbot, V., Theotokatos, G., Bujorianu, M. L., Boulougouris, E., & Vassalos, D. (2018). Vulnerabilities and safety assurance methods in Cyber-Physical Systems: A comprehensive review. *Reliability Engineering & System Safety*.

Books, H. (1997). Successful health and safety management. *HS (G)*.

Dekker, S. (2014). *The field guide to understanding'human error'*: Ashgate Publishing, Ltd.

[†] www.nexus-project.eu

DNVGL. (2015). Dynamic positioning vessel design philosophy guidelines. Recommended practice (DNVGL-RP-E306). In.

Hollnagel, E. (2017). *The ETTO principle: efficiency-thoroughness trade-off: why things that go right sometimes go wrong*: CRC Press.

IMCA. (2015). International Guidelines for The Safe Operation of Dynamically Positioned Offshore Supply Vessels (182 MSF Rev. 2). In.

IMO. (1994). Guidelines for vessels with dynamic positioning systems (IMO MSC Circular 645). In. London.

Lee, J. D. (2008). Review of a pivotal Human Factors article:"Humans and automation: use, misuse, disuse, abuse". *Human Factors, 50*(3), 404-410.

Leveson, N. (2011). *Engineering a safer world: Systems thinking applied to safety*: MIT press.

Leveson, N. (2015). A systems approach to risk management through leading safety indicators. *Reliability Engineering & System Safety, 136*, 17-34.

Leveson, N., & Thomas, J. (2018). *STPA Handbook*. Retrieved from http://psas.scripts.mit.edu/home/materials/

Leveson, N. G. (1995). Safeware. *System Safety and Computers. Addison Wesley*.

Leveson, N. G. (2011). Applying systems thinking to analyze and learn from events. *Safety Science, 49*(1), 55-64.

Perrow, C. (1984). Normal accidents: Living with high risk systems. In: New York: Basic Books.

Rae, A., McDermid, J., & Alexander, R. (2012). The science and superstition of quantitative risk assessment. *Journal of Systems Safety, 48*(4), 28.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science, 27*(2), 183-213.

Rokseth, B., Utne, I. B., & Vinnem, J. E. (2017). A systems approach to risk analysis of maritime operations. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 231*(1), 53-68.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics, 2*, 1926-1943.

Skjong, R., & Wentworth, B. H. (2001). *Expert judgment and risk perception.* Paper presented at the The Eleventh International Offshore and Polar Engineering Conference.

Tiusanen, R. (2017). Qualitative Risk Analysis. *Handbook of Safety Principles*, 463-492.

Young, R. R. (2004). *The requirements engineering handbook*: Artech House.

Towards STAMP approach based protection of Underwater Cultural Heritage

Robert Aps[1*], **Liisi Lees**[1], **Kristjan Herkül**[1], **Maili Roio**[2], **Olev Tõnismaa**[3]
[1] University of Tartu, Estonian Marine Institute, Estonia
[2] National Heritage Board of Estonia, Tallinn, Estonia
[3] Tallinn University of Technology, Estonian Maritime Academy, Tallinn, Estonia

ABSTRACT

According to the UN Law of the Sea Convention (LOSC), states have the duty to protect objects of an archaeological and historical nature found at sea and shall cooperate for this purpose. The 2001 UNESCO Convention on the Protection of the Underwater Cultural Heritage stipulates that in-situ preservation of underwater cultural heritage (i.e. on the seabed) must be considered as the first and preferred option before allowing or engaging in any activities directed at this heritage. To prevent incidental damage the State Party shall use the best practicable means at its disposal to prevent or mitigate any adverse effects that might arise from activities under its jurisdiction incidentally affecting underwater cultural heritage. A Systems-Theoretic Accident Modelling and Processes (STAMP) approach to operational safety management considers accident occurrence as the result of a lack of, or inadequate enforcement of, constraints imposed on the system design and operations at various system levels. The objective of this study in progress is to apply the STAMP based Systems-Theoretic Process Analysis (STPA) to identify the system level hazards and potentially unsafe ship anchoring control actions incidentally affecting and damaging underwater cultural heritage objects in the Estonian national Vessel Traffic Service (VTS) Centre sea area in the Gulf of Finland, Baltic Sea. The physical damage to underwater monuments and heritage conservation areas caused by anchoring of ship is identified as an accident (an undesired and unplanned loss event) and the legal protection and preservation restrictions applicable to underwater monuments and the protected zone thereof are considered to be the underwater cultural heritage protection and preservation constraints to be enforced. The critical role of VTS in effective hazard control actions and the enforcement of preventive constraints in real time is identified.

Keywords: Systems-Theoretic Accident Modelling and Processes; Systems-Theoretic Process Analysis; Underwater Cultural Heritage; ship anchoring.

1. INTRODUCTION

According to the UN Law of the Sea Convention (LOSC) states have the duty to protect objects of an archaeological and historical nature found at sea and shall cooperate for this purpose (UN, 1982). The 2001 UNESCO Convention on the Protection of the Underwater Cultural Heritage (UNESCO, 2001) stipulates that in-situ preservation of underwater cultural heritage (i.e. on the seabed) must be considered as the first and preferred option before allowing or engaging in any activities directed at this heritage. It is stated also that underwater cultural heritage faces a wide array of threats and negative impacts that endanger its preservation and therefore the protection of underwater cultural heritage is at the heart of this Convention, together with public enjoyment and the fight against commercial exploitation. It is specified further that activities incidentally affecting underwater cultural heritage, despite not having underwater cultural heritage as their primary object or one of their objects, may physically disturb or otherwise damage underwater

---

[*] Corresponding author: phone: +372 5062597, email address: robert.aps@ut.ee

cultural heritage. The State Party shall use the best practicable means at its disposal to prevent or mitigate any adverse effects that might arise from activities under its jurisdiction incidentally affecting underwater cultural heritage.

Referring to Estonian Heritage Conservation Act in force (EHCA, 2019) “A monument is a movable or immovable, a part thereof, a body of things or an integral group of structures under state protection which is of historical, archaeological, ethnographic, urban developmental, architectural, artistic or scientific value or of value in terms of religious history or of other cultural value and due to which it is designated as a monument pursuant to the procedure provided for in this Act”. It is stated further, that the underwater monuments can be things or bodies of things specified by this Act which are located in internal and transboundary water bodies, inland and territorial seas and exclusive economic zones. It is prohibited to destroy or damage monuments and as additional restrictions applicable to underwater monuments and the protected zone thereof it is prohibited to anchor, trawl, dredge and dump solid substances within underwater monuments and the protected zones thereof.

The Systems-Theoretic Accident Model and Processes (STAMP) approach considers safety as emergent property of the system, arising from the interaction of system components within a given environment and the accident occurrence as the result of a lack of, or inadequate enforcement of, constraints imposed on the system design and operations at various system levels (Leveson, 2011). In STAMP the safety is viewed as a control problem, and safety is managed by a control structure embedded in an adaptive socio-technical system while the system itself is viewed as interrelated components that are kept in a state of dynamic equilibrium by feedback loops of information and control (Leveson, 2004). Thus, the basic concepts in STAMP are constraints, control loops, process models and levels of control, and the safety management is defined as a continuous control task to impose the constraints necessary to limit system behavior to safe changes and adaptations.

The STAMP based Systems-Theoretic Process Analysis (STPA) (Leveson, 2011; Thomas, 2012) is a powerful new hazard analysis method designed to go beyond traditional safety techniques has been successfully applied e.g. to space engineering applications (Ishimatsu, Leveson, Thomas, Fleming, Katahira, Miyamoto, Ujiie, Nakao, & Hoshino, (2014) as well as to analysis of maritime traffic safety in the Gulf of Finland (Aps, Fetissov, Goerlandt, Kujala, & Piel, 2017). However, the STAMP approach based protection of Underwater Cultural Heritage has attracted less attention so far.

This study is a part of the INTERREG BSR project “Baltic Sea Region Integrated Maritime Cultural Heritage Management (BalticRIM)". The aim of this study in progress was to apply the STAMP based STPA methodology to the underwater cultural heritage management domain. The objective was to identify the system level hazards and potentially unsafe ship anchoring control actions incidentally affecting and damaging underwater cultural heritage objects in the Estonian part of the BalticRIM Tallinn-Helsinki pilot area covered by national Vessel Traffic Service (VTS) in the Gulf of Finland, Baltic Sea.

2. STUDY AREA

According to IMO (2006) “The mandatory ship reporting system in the Gulf of Finland covers the international waters in the Gulf of Finland. In addition, Estonia and Finland have implemented mandatory ship reporting systems to their national water areas outside VTS areas. These reporting systems provide the same services and make the same requirements to shipping as the system operating in the international waters. The mandatory ship reporting system and the Estonian and Finnish national mandatory ship reporting systems are together referred as the GOFREP and their area of coverage respectively as the GOFREP area” (Figure 1).

Facilitation of exchange of information between the ship station and the shore station aiming at supporting the safe navigation and the protection of the marine environment is seen as the primary objective of the GOFREP system. The GOFREP/VTS Center operator is able to observe the controlled maritime traffic process through the radar and Automatic Identification System (AIS) surveillance of traffic and to actuate the process if the ship under control proceed against ship anchoring adjustment to a safe level appropriate to protection and preservation of underwater cultural heritage requirements.

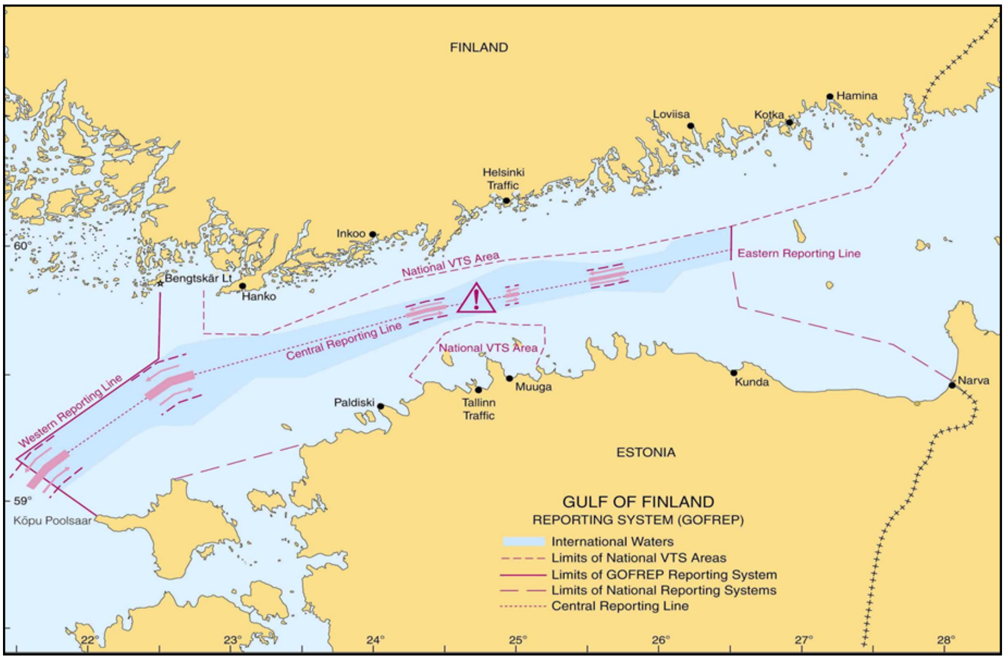


Figure 1. The mandatory ship reporting system in the Gulf of Finland (Baltic Sea)
(Source: Estonian Maritime Administration)

The GOFREP maritime traffic control system is jointly managed by the Finnish Transport Agency, Estonian Maritime Administration and the Federal Agency for Maritime and River Transport of Russian Federation and is based on the activities of GOFREP Traffic Centers of Estonia (Tallinn Traffic), Finland (Helsinki Traffic) and the Russian Federation VTMIS Centre in Petrodvorets (Saint Petersburg Traffic).

The BalticRIM Tallinn-Helsinki pilot sea area is situated in the central part of the Gulf of Finland and the Estonian part of this pilot sea area extends from shoreline to the outer border of the Estonian exclusive economic zone (Figure 2).

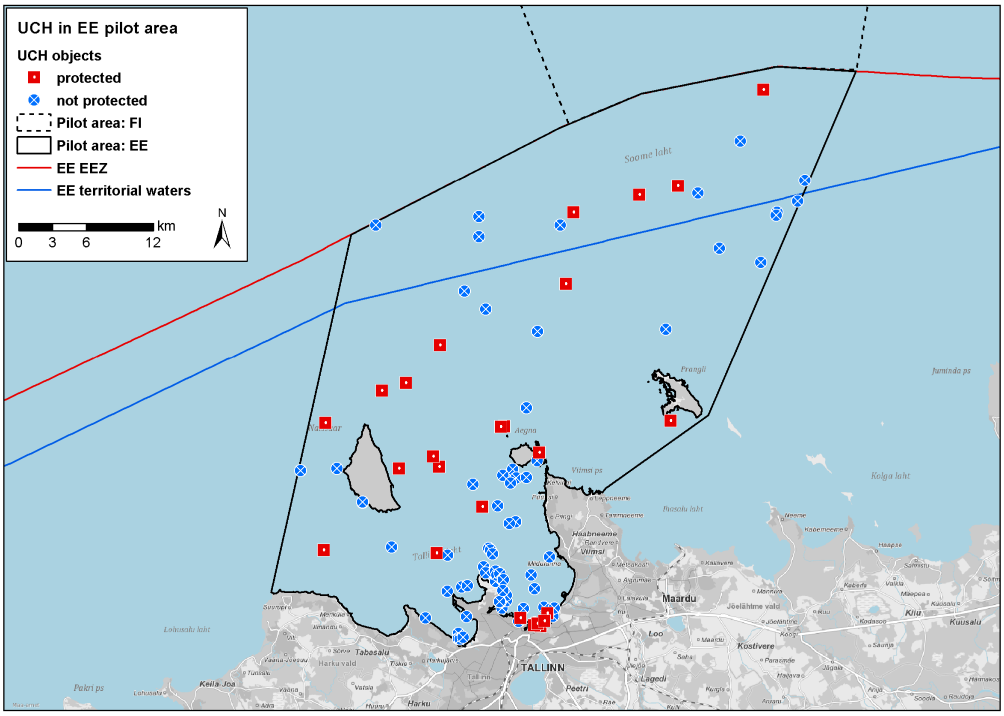


Figure 2. Locations of underwater cultural heritage (UCH) objects in the BalticRIM Tallinn-Helsinki pilot sea area (Source: Estonian National Registry of Cultural Monuments and Estonian Maritime Administration)

Underwater cultural heritage objects in the BalticRIM Tallinn-Helsinki pilot sea area and the high shipping intensity AIS pattern in the background are presented in the Figure 3.

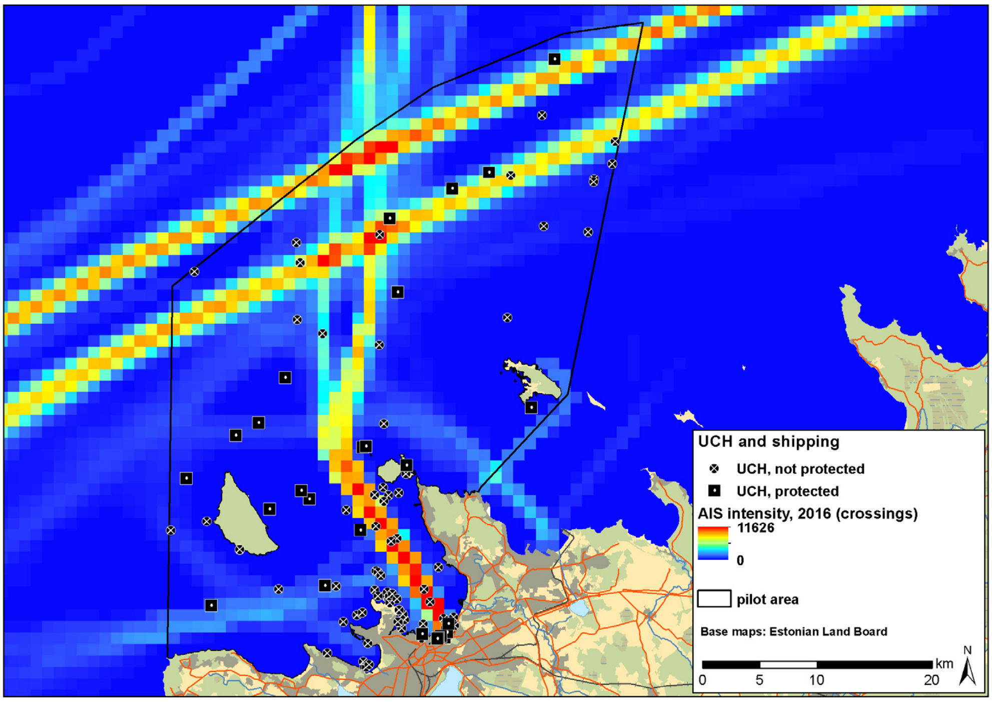


Figure 3. Underwater cultural heritage (UCH) objects in the BalticRIM Tallinn-Helsinki pilot sea area and the high shipping intensity AIS pattern in the background
(Source: Estonian National Registry of Cultural Monuments and Estonian Maritime Administration, HELCOM Map and Data Service)

The BalticRIM Tallinn-Helsinki pilot sea area is characterized by high shipping intensity (Figure 3). The ship anchoring within this shipping intensive area is one of the biggest threats to underwater monuments and heritage conservation areas and therefore the analysis focuses on that threat.

At the same time, to enable effective hazard control and the enforcement of preventive constraints in real time, it is established (IMO, 2003) that on receipt of a position message, the GOFREP/VTS operators are determining the relationship between the ship position and the information supplied by the position-fixing equipment available to them while the information on course and speed is helping operators to identify one ship among a group of ships. This is achieved automatically if the Automatic Identification System (AIS) transponder is used. If necessary, individual information can be provided to a ship, particularly in relation to positioning and navigational assistance or local conditions and if a ship needs to anchor due to breakdown or emergency the operator can recommend suitable anchorage in the area.

2. MARITIME NAVIGATION SAFETY MANAGEMENT IN THE GULF OF FINLAND

The levels of hierarchical structure of maritime navigation safety management in the Gulf of Finland from European to ship on-board level (Figure 4) are connected by communication channels, and referring to Leveson (2011) "… a downward reference channel is providing the information necessary to impose safety constraints on the level below and an upward measuring channel to provide feedback about how effectively the constraints are being satisfied".

The SafeSeaNet is functioning at the European level as the maritime information and exchange system established to facilitate the exchange of information in an electronic format between EU Member States and to provide the Commission with relevant information in accordance with Community legislation.

The GOFREP/VTS Centers represent the onshore level of maritime traffic safety management and communication in the Gulf of Finland. According to IMO (2003), the functions of GOFREP/VTS Centers are performed through a combination of 1) radar and Automatic

Identification System (AIS) surveillance of traffic and navigational marks in the Ship Reporting System (SRS) sea area with particular scrutiny of the development of conflicts in ship traffic, 2) radio communication, and 3) the maintenance of direct and separate communication links between the GOFREP/VTS Centers for the exchange, updating and co-ordination of information. The system is capable of providing an automatic alarm to identify any track that strays into the unauthorized area.

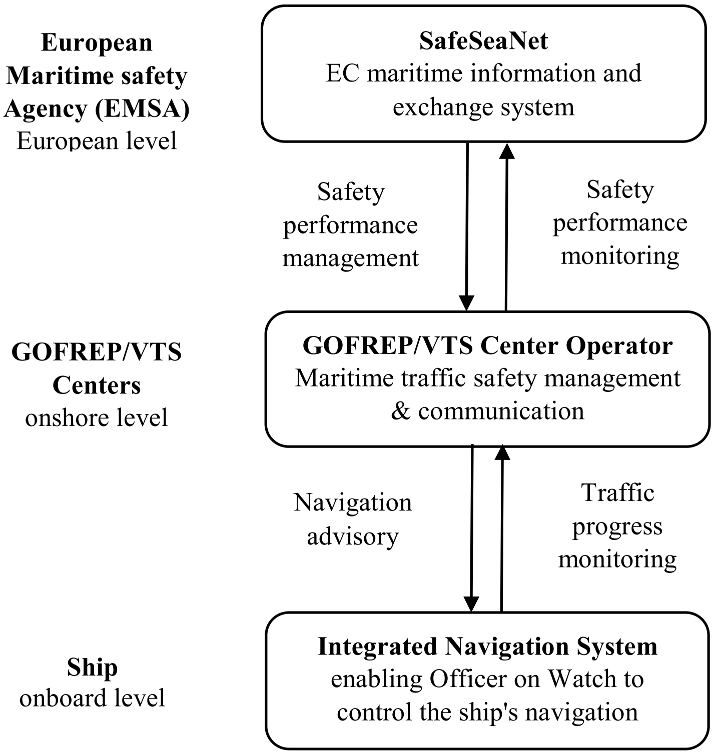

European
Maritime safety
Agency (EMSA)
European level
SafeSeaNet
EC maritime information and exchange system
Safety performance management
Safety performance monitoring
GOFREP/VTS
Centers
onshore level
GOFREP/VTS Center Operator
Maritime traffic safety management & communication
Navigation advisory
Traffic progress monitoring
Ship
onboard level
Integrated Navigation System
enabling Officer on Watch to control the ship's navigation

Figure 4. Hierarchical structure of maritime navigation safety management from European to ship onboard level (modified from Leveson, 2011)

The ship onboard level is characterized by the Integrated Navigation System that provides 'added value' to the functions and information needed by the Officer on Watch to plan, monitor or control the progress of the ship. However, as argued by House (2007) “Shipping the world over is notorious for experiencing the unusual and the unexpected. In most cases if and when routine practice goes wrong, the weather is usually a key element which influences the cause and very often the outcome. The other variable is often the human element which can work for, or against, the wellbeing of the ship”. Additionally, ship-related hazards are associated with ship-specific equipment or operations.

3. STPA HAZARD ANALYSIS

Referring to Leveson (2011) “Hazard analysis can be described as ‘investigating an accident before it occurs’. The goal is to identify potential causes of accident, that is, scenarios that can lead to losses, so they can be eliminated or controlled in design or operations before damage occurs”.

As stated by Thomas (2012) “The first step in STPA is to identify the potentially unsafe control actions for the specific system being considered. These unsafe control actions are used to create safety requirements and constraints on the behavior of both the system and its components. Additional analysis can then be performed to identify the detailed scenarios leading to the violation of the safety constraints. As in any hazard analysis, these scenarios are then used to control or mitigate the hazards in the system design”. It is added that before beginning an STPA hazard analysis, potential accidents and related system-level hazards are identified along with the corresponding system safety constraints that must be controlled.

It is further specified (Thomas, 2012) - while potential unsafe control actions are identified in the first step of STPA, the second step examines their control loops to identify causal factors for each unsafe control action, i.e., the scenarios for causing the hazard.

In this study the potential physical damage to underwater monuments and heritage conservation areas caused by ship anchoring is identified as an accident (an undesired and unplanned loss event). The legal protection and preservation restrictions stipulated by Estonian Heritage Conservation Act (EHCA, 2019) applicable to underwater monuments and the protected zone thereof, are considered to be the underwater cultural heritage protection and preservation constraints to be enforced.

3.1. System high level hazards and constraints

According to Leveson (2011), a hazard is a system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to an accident (loss). It is added further that hazards may be defined in terms of conditions or in terms of events as long as one of these choices is used consistently and the only difference is that the events are limited in time while the conditions caused by the event persist over time until another event occurs that changes the prevailing conditions.

Underwater cultural heritage protection and preservation related high level hazard and ship anchoring constraints that are to be enforced are presented in Table 1.

Table 1. Underwater cultural heritage protection and preservation related high level hazard and ship anchoring constraints

Underwater cultural heritage protection and preservation related high level hazard	Underwater cultural heritage protection and preservation related ship anchoring constraints
Controlled ship violate underwater cultural heritage protection and preservation related anchoring requirements	According to Estonian Heritage Conservation Act (2019) it is prohibited to anchor within underwater monuments and the protected zones thereof. Every ship shall at all times use all available means appropriate to the prevailing circumstances and conditions to avoid the potential physical damage to underwater monuments and the protected zone thereof caused by ship anchoring

3.2. Potentially unsafe ship anchoring control actions

Referring to Thomas (2012) “STAMP is based on the observation that there are four types of hazardous control actions that need to be eliminated or controlled to prevent accidents:

1. A control action required for safety is not provided or is not followed
2. An unsafe control action is provided that leads to a hazard
3. A potentially safe control action is provided too late, too early, or out of sequence
4. A safe control action is stopped too soon or applied too long”.

In the context of this study in progress, when situations occur of controlled ship violating underwater cultural heritage protection and preservation related safe anchoring requirements, the control action is required on ship anchoring adjustment to a safe level appropriate to protection and preservation of underwater cultural heritage.

When the required action on ship anchoring adjustment to a safe level appropriate to preservation and protection of underwater cultural heritage is not provided, is provided incorrectly or is provided too late, the system is led to a hazardous state defined as a violation of underwater cultural heritage protection and preservation related safe anchoring requirements.

As the first step of STPA, any potentially unsafe control actions on ship anchoring adjustments to a safe level appropriate to protection and preservation of underwater cultural heritage are identified based on interviews of maritime navigation professionals and their relevant discussions and presented in Table 2.

Table 2. Potentially unsafe control actions on ship anchoring adjustment to a safe level appropriate to protection and preservation of underwater cultural heritage

Control action required	Action required but not provided	Action provided unsafe	Action provided			Stopped too soon
			Too early	Too late	Out of sequence	
Control action on ship anchoring adjustment to a safe level appropriate to protection and preservation of underwater cultural heritage	Hazardous state – ship anchoring is not adjusted to a safe level appropriate to protection and preservation of underwater cultural heritage	Hazardous state – ship anchoring is not adjusted properly to a safe level appropriate to protection and preservation of underwater cultural heritage	N/A	Hazardous state – ship anchoring is not adjusted timely to a safe level appropriate to protection and preservation of underwater cultural heritage	N/A	N/A

3.3. Scenario leading to potentially unsafe ship anchoring control actions

The second step of STPA hazard analysis is performed on a STAMP-Mar standard control loop of the integrated navigation system operated at the ship onboard level (Figure 5). The aim is to identify the causal factors for potentially hazardous control actions on ship anchoring adjustment to a safe level appropriate to protection and preservation of underwater cultural heritage. Analysis is based on interviews of experts - maritime navigation professionals and their topic focused discussions.

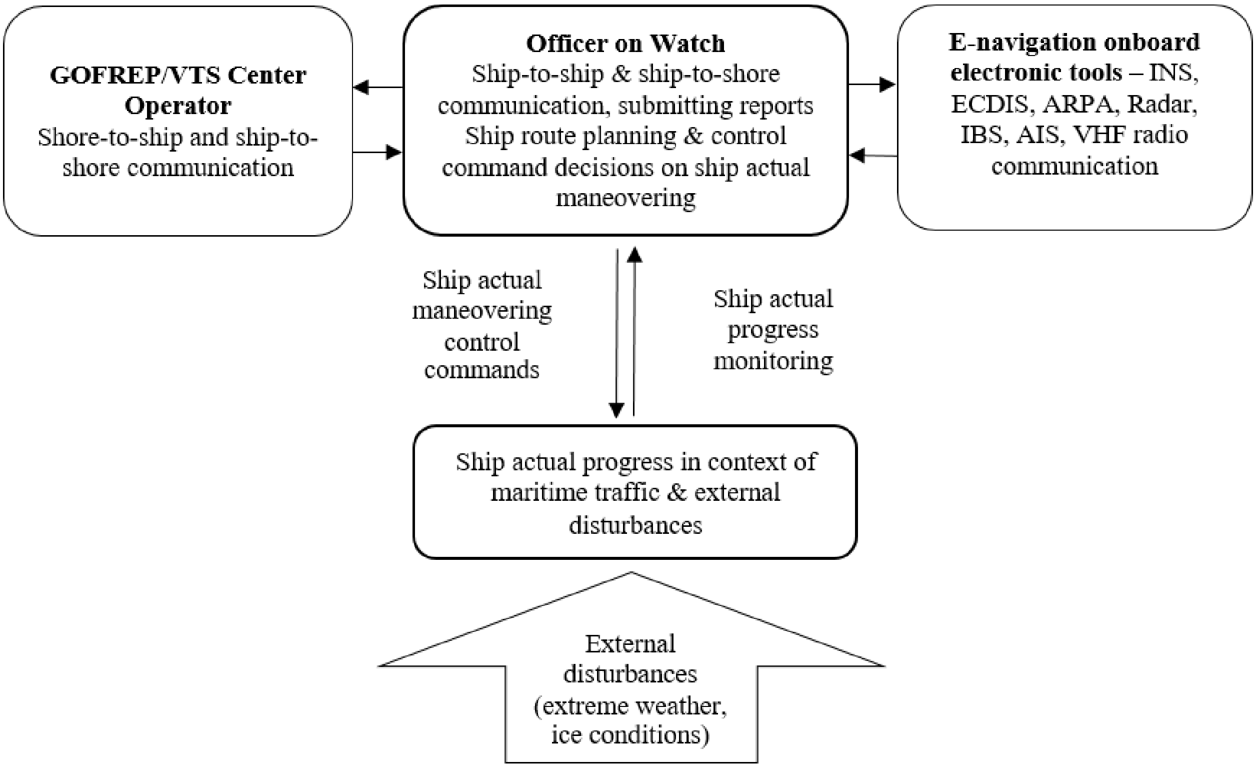

GOFREP/VTS Center Operator
Shore-to-ship and ship-to-shore communication
Officer on Watch
Ship-to-ship & ship-to-shore communication, submitting reports
Ship route planning & control command decisions on ship actual maneovering
E-navigation onboard electronic tools – INS, ECDIS, ARPA, Radar, IBS, AIS, VHF radio communication
Ship actual maneovering control commands
Ship actual progress monitoring
Ship actual progress in context of maritime traffic & external disturbances
External disturbances (extreme weather, ice conditions)

Figure 5. The STAMP-Mar standard control loop of the integrated navigation system operated at the ship onboard level (modified from Leveson, 2011)

As a result, experts have suggested that incomplete awareness of the situation by the Officer on Watch caused by malfunction of one or more e-navigation onboard tools (e.g. satellite navigation system, ARPA, radar equipment, AIS) should be considered an important causal factor leading to potentially hazardous control actions on ship anchoring adjustment to a safe level appropriate to protection and preservation of underwater cultural heritage.

3.4. Enforcement of safety constraints

The STAMP-Mar standard control loop of the integrated navigation system operated at the at the ship onboard level (Figure 5) has been verified and discussed by experts to ensure that the safety constraints for identified scenarios (the incomplete awareness of the situation by the Officer on Watch due to malfunction of one or more e-navigation on-board tools) can truly be enforced in system operations. Efficient ship-to-shore and shore-to-ship communication is recognized as a fundamentally important control factor to update the awareness of the Officer on Watch effectively and in real time.

With respect to enforcement of underwater cultural heritage protection and preservation related anchoring safety constraints, and referring to (IMO, 1997), the WTS is "… a service implemented by a Competent Authority, designed to improve the safety and efficiency of vessel traffic and to protect the environment. The service should have the capability to interact with the traffic and to respond to traffic situations developing in the VTS area". It is added that VTS should comprise at least an information service to ensure that essential information becomes available in time for on-board navigational decision-making and to monitor its effects. It is specified further that the information service is provided by broadcasting information at fixed times and intervals or when deemed necessary by the VTS or at the request of a vessel, and may include for example reports on the position, identity and intentions of other traffic, waterway conditions, weather, hazards, or any other factors that may influence the vessels' transit. The navigational assistance service is especially important in difficult navigational or meteorological circumstances or in case of defects or deficiencies being normally rendered at the request of a vessel or by the VTS when deemed necessary.

In accordance with the IMO Guidelines and Criteria for Ship Reporting systems (IMO, 1994) the communication between a VTS authority and a participating ship should be conducted and should be limited to information essential to achieve the objectives of the VTS. At that, the IMO Standard Marine Communication Phrases (IMO, 2001) should be used when practicable. In addition, any VTS message directed to a ship or ships should be clear whether the message contains information, advice, a warning, or an instruction. It is suggested (IALA, 2012) that in order to further facilitate shore-to-ship and ship-to-shore communication in a VTS environment, one of the following eight message markers should be used to increase the likelihood that the purpose of the message is properly understood (information, warning, advice, instruction, question, answer, request and intention) leaving it at the discretion of the shore personnel or the ship officer whether to use one of the message markers and, if so, which marker is applicable to the situation.

Furthermore, referring to IALA (2012) the message marker 'Warning' is used to convey potentially dangerous situations or observe developing situations. The contents of a warning message should be assessed immediately in conjunction with any additional information that may not be available to the VTS Center and corrective action taken when necessary.

A fundamental principle of VTS communications (IMO, 1997) is that when the VTS is authorized to issue 'Instructions' to ship, "… these instructions should be result-oriented only, leaving the details of execution, such as course to be steered or engine manoeuvres to be executed, to the master or pilot on board the vessel. Care should be taken that VTS operations do not encroach upon the master's responsibility for safe navigation, or disturb the traditional relationship between master and pilot". The message marker 'Instruction' conveys that the message is a directive given by the VTS Center under the provisions of a statutory regulation and the sender must have delegated authority to send such a message (IALA, 2012). For example, with aim to support an action on ship anchoring adjustment to a safe level appropriate to protection and preservation of underwater cultural heritage, the 'Instruction' messages like 'Anchoring is prohibited as you are in an area of underwater cultural heritage' should be issued to the ship concerned.

4. CONCLUSIONS

The potential physical damage to underwater monuments and heritage conservation areas caused by ship anchoring is identified as an accident (an undesired and unplanned loss event). The legal preservation and protection restrictions stipulated by Estonian Heritage Conservation Act, in force and applicable to underwater monuments and the protected zone thereof, are considered to be the underwater cultural heritage preservation and protection constraints to be enforced.

The STPA hazard analysis is performed in order to identify the causal factors and scenarios for potentially hazardous ship anchoring control actions based on interviews of experts and their relevant discussions. As a result, the incomplete awareness of the situation by the Officer on Watch due to malfunction of one or more e-navigation on-board tools was identified as the potential hazardous scenario leading to anchoring within underwater monuments and the protected zones thereof. The critical role of VTS Centre in effective hazard control actions and the enforcement of preventive constraints in real time is identified.

The GOFREP/VTS Center operator is able to observe the controlled maritime traffic process through the radar and Automatic Identification System (AIS) surveillance of traffic. The operator is also able to actuate the process if the ship under control proceed against ship anchoring adjustment to a safe level appropriate to protection and preservation of underwater cultural heritage requirements, by issuing the 'Instruction' messages like 'Anchoring is prohibited as you are in an area of underwater cultural heritage' to the ship concerned.

ACKNOWLEDGEMENTS

This study is supported by European Regional Development Fund, INTERREG Baltic Sea Region project BalticRIM "Baltic Sea Region Integrated Maritime Cultural Heritage Management".

The authors thank the Working Group for Marine Planning and Coastal Zone Management (WGMPCZM) of the International Council for the Exploration of the Sea (ICES) for facilitating this research.

REFERENCES

Aps, R., Fetissov, M., Goerlandt, F., Kujala, P., Piel, A. (2017). Systems-Theoretic Process Analysis of maritime traffic safety management in the Gulf of Finland (Baltic Sea). *Procedia Engineering,* 179, Elsevier, 2-12.

EHCA (2019). Estonian Heritage Conservation Act. Retrieved from https://www.riigiteataja.ee/en/eli/ee/Riigikogu/act/504042019007/consolide

House, D.J. (2007). *Ship handling – theory and practice*. Elsevier.

IALA (2012). Vessel traffic services manual, Edition 5, IALA-AISM. Retrieved from https://www.pmo.ir/pso_content/media/files/2013/1/22176.pdf

IMO (2006). Mandatory Ship Reporting Systems, SN.1/Circ.258, Retrieved from https://www.transportstyrelsen.se/contentassets/a498840125d6473e8046aec0c261633d/258.pdf

IMO (2003). Mandatory ship reporting systems, SN/Circ.225. Retrieved from https://www.transportstyrelsen.se/contentassets/a498840125d6473e8046aec0c261633d/225.pdf

IMO (2001). Standard marine communication phrases, Resolution A.918 (22). Retrieved from https://www.transportstyrelsen.se/contentassets/a19ffd185d54440fbea6581bde25ee3b/918.pdf

IMO (1997). Guidelines for vessel traffic services, Resolution A.857 (20). Retrieved from https://www.transportstyrelsen.se/contentassets/a19ffd185d54440fbea6581bde25ee3b/857.pdf

IMO (1994). Guidelines and criteria for ship reporting systems, MSC.43 (64). Retrieved from https://puc.overheid.nl/nsi/doc/PUC_1376_14/3/

Ishimatsu, T., Leveson, N., Thomas, J., Fleming, C., Katahira, M., Miyamoto, Y., Ujiie, R., Nakao, H., Hoshino, N. (2014). Hazard Analysis of Complex Spacecraft Using Systems-Theoretic Process Analysis, *Journal of Spacecraft and Rockets*, 51(2), 509–522.

Leveson, N. (2004). A New Accident Model for Engineering Safer Systems. *Safety Science*, 42, 237-270.

Leveson, N. (2011). *Engineering a safer world - Systems theory applied to safety.* Cambridge, MA, MIT Press.

Thomas, J.P. (2012). Extending and Automating a Systems-Theoretic Hazard Analysis for Requirements Generation and Analysis. *Sandia Report SAND2012-4080*. Sandia National Laboratories.

UN (1982). United Nations Convention on the Law of the Sea (LOSC). Retrieved from http://www.un.org/Depts/los/convention_agreements/texts/unclos/unclos_e.pdf

UNESCO (2001). Convention on the protection of the underwater cultural heritage. Retrieved from http://www.unesco.org/new/en/culture/themes/underwater-cultural-heritage/2001- convention/

System-theoretic process analysis for safety analysis of cooperative material handling machinery – Concept and initial experiences

Tommi Kivelä[1,*] and Kai Furmans[1]
[1] Institute for Material Handling and Logistics, Karlsruhe Institute of Technology, Germany

ABSTRACT

In material handling and logistics, there’s a trend towards increasingly adaptable and flexible approaches on all system levels: from the supply chain and logistic network level down to the factory and warehouse floors. Recent examples of increasingly flexible material handling technologies on the floor level are autonomously navigating automated guided vehicles (AGVs) and plug-and-work material handling systems, the first allowing adaptable material flow systems with minimal fixed infrastructure, the latter allowing the user to easily re-configure steady conveyor systems on demand. In the field of safety engineering, there has recently been research towards safety assurance of open adaptive systems (OAS) with frameworks such as runtime certification as potential enabler for these novel systems. In this work, we seek to combine recent concepts from the safety engineering community with traditional and advanced technologies from the area of material handling machinery to enable the next step in operational flexibility in this application area. We suggest potential application use cases which would be enabled by the use of dynamic safety contracts: safely cooperating material handling machinery. Compared to machinery with traditional, fixed interfaces, the machine-to-machine cooperation will increase the complexity of the required safety-related control systems and software, which will in turn require new approaches for the risk assessment and safety engineering of these types of systems. We suggest the use of STPA for safety-driven design of cooperative material handling machinery. We discuss one novel application concept, AGV-Storage crane cooperative handover, in detail and present initial results of STPA analysis for the application.

Keywords: STPA; Machinery safety; Material handling; Industry 4.0; Runtime certification.

1. INTRODUCTION

In material handling and logistics, there’s a trend towards increasingly adaptable and flexible approaches on all system levels: from the supply chain and logistic network level down to the factory and warehouse floors (Delfmann, Hompel, Kersten, Schmidt, & Stölzle, 2018). The increasing adaptability and flexibility on the machinery level is driven by increasingly autonomous and networked machines employing decentralised control (Delfmann et al., 2018; Furmans, Schonung, & Gue, 2010). For increasingly complicated machinery applications, new approaches are needed for risk assessment and engineering of safe systems. In this work, we suggest an approach using dynamic safety contracts to enable safe cooperation between material handling machinery to enable more flexibility in material flow systems.

The suggested automated cooperation among machines increases the system complexity and brings challenges to the risk assessment performed during the design phase of the machinery. We suggest using System-Theoretic Process Analysis (STPA), a risk analysis method based on the System-Theoretic Accident Model and Processes (STAMP), developed by Leveson (2011).

The work is structured as follows. In section 2 we discuss related work: recent advances in the fields of material handling, such as plug & play material handling and the move towards decentralised control, and in the field of safety engineering, such as recent advances towards safety of open adaptive systems. We additionally provide a brief overview of STAMP and STPA. In

---

[*] Corresponding author: +49 721 608-48645, tommi.kivelae@kit.edu

section 3 we discuss the current regulation and state of the art in safety engineering in the machinery industry, mainly from the point of view of the European regulatory framework. In section 4 we introduce the concept of cooperative material handling machinery, and discuss possible new use cases enabled by dynamic safety contracts. Finally, in section 5 we utilise STPA to analyse one example application with cooperative material handling machinery. Last, we discuss our initial experiences with the method and the results of the analysis as well as future work and research directions.

2. RELATED WORK

2.1 Recent approaches in material handling

Some recent trends in the area of material handling machinery have been increasing autonomy and decentralisation of control, enabling so called plug-and-play or plug-and-work material handling systems. Increasing autonomy in material handling machinery is best exemplified by the developments in automated guided vehicle (AGV) technology. Modern vehicles utilise, for example, laser scanners for autonomous navigation, with the help of natural landmarks and additional infrastructure (reflectors as artificial landmarks) (Ullrich & Kachur, 2015) or most recently only based on natural landmarks, as the AGVs developed in the KARIS PRO-project (Project Consortium KARIS PRO, 2017). This autonomy is enabled by safety technology integrated in each vehicle: laser scanners for protection fields, connected to safety controllers which bring the vehicle to a stop if humans or other objects are detected in the protection fields. Autonomous vehicles enable flexible material handling operations, which can adapt to changing layouts, for example, in a dynamic production environment.

Furmans, Schonung and Gue (2010) have suggested design patterns for flexible future material handling systems: material handling systems should consist of highly independent modules (modularity), which contain all functions necessary to perform their tasks (function integration). Additionally, the actions of the modules are controlled by their own controllers (decentralised control), with adjacent modules freely exchanging information and goods (interaction). This should be possible through the use of standardised physical and information interfaces. Some examples of such plug-and-work material handling systems are the FlexConveyor (Mayer, 2009) and the GridSorter (Seibold, 2016). As another example of decentralised control, the KARIS PRO AGV implements the task management system through decentralised decision making (Colling, Ibrahimpasic, Trenkle, & Furmans, 2016), removing the need of a centralised task management system.

2.2 Safety of Open Adaptive Systems

In safety engineering, recent research has focused on open adaptive systems, driven by, for example, challenges related to cooperative autonomous driving. One framework for safety assurance in open adaptive systems was the concept of runtime certification, first suggested by Rushby (2008). Rushby (2008) suggested that parts of the traditional certification process could be automated and transferred to “runtime”. Compliance to standards is replaced by assurance cases, constructing explicit goals, evidences and arguments. These assurance cases could be formally verified at runtime (runtime verification) thus “certifying” the system. In later work Rushby (2016) discusses a possible medical application based on self-integration of safety-related systems.

For example Trapp and Schneider (2014) have built on top of Rushby’s suggestions and they suggest different possible approaches, where the traditional safety engineering is increasingly transferred to runtime for increased flexibility at the cost of increased complexity. The simplest case is the so called safety certificate at runtime-approach, where subsystems check for assumptions and demands for safe integration. The integration scenarios are pre-engineered at design time, while the exact runtime integration partners are not known. Schneider (2014) studied this approach based on the concept of conditional certificates, ConSerts, for tractor implement automation (TIA). The ConSert-approach was used as a basis for dynamic safety contracts for automotive cooperative applications by Müller and Liggesmeyer (2016) and for cooperative medical applications by Leite, Schneider, and Adler (2018).

There have been further advances such as the approach of Calineschu et al. (2018) for adapting assurance cases at runtime, but here we focus on the simplest case of runtime certification, where the integration scenario is pre-designed. The machines have pre-defined interfaces to enable the dynamic safety contract for a specific application scenario.

A similar approach is studied currently in the SmartFactoryKL project (Popper et al., 2018), however the focus is on modular machine configurations (automated integration of machine modules to form a complete machine), whereas our focus here is on dynamic safety contracts between complete machines.

2.3 STAMP and STPA

The accident causality model STAMP, suggested by Leveson (2011), is based on systems theory. Safety is considered an emergent system property and is treated as a dynamic control problem instead of a failure prevention problem. The accident causality model has been used as a basis to create a new risk analysis method, STPA. In STPA the studied system is modelled as a hierarchical control structure, which is used as a basis to find unsafe control actions and causal factors against which further design control or other mitigation measures can be implemented. Due to the background in systems theory, STPA is better suited to detect complex causal chains in modern software-intensive systems. (Leveson, 2011; Leveson & Thomas, 2018)

STPA has been studied extensively in other industries such as defense (Leveson, 2011), aviation (Fleming & Leveson, 2014; Fleming, Spencer, Thomas, Leveson, & Wilkinson, 2013) and automotive (Abdulkhaleq, 2017). So far there seems to be little published work about using STPA for machinery systems, though a robotics example is provided by Leveson (2011).

3. MACHINERY SAFETY ENGINEERING

Machine products sold on the European market must adhere to the essential health and safety requirements originating from the Machinery Directive (European Parliament and The Council of the European Union, 2006). Typically machinery manufacturers follow harmonized safety standards, such as the ISO 12100 (2010), to ensure conformance to the Machinery Directive. A typical machine lifecycle model and the risk assessment & reduction process according to ISO 12100 (2010) is illustrated in Figure 1.

When a new machine is designed, the risks related to hazards arising from the machine are to be assessed and reduced to an acceptable level. The focus of the Machinery Directive (European Parliament and The Council of the European Union, 2006) and the machinery safety standardisation is the safety of the machine operators and personnel in the vicinity of the machinery. According to ISO 12100 (2010), risk is considered to be the combination of the probability of occurrence of harm and the severity of that harm, harm being physical injury or damage to health, which might occur due to a hazard. The risk reduction can be achieved through inherently safe design, technical protective measures or lastly through information for use. Technical protective measures are designed according to the relevant functional safety standards, with additional information, such as the required rigor of implementation, defined in product standards. If the machine or application is not yet covered by a product standard, risk graphs in, for example, ISO 13849 (2015) can be used for risk estimation. If the safety-related functions are used as a risk reduction measure, the functional safety standards typically require strict separation of the safety-related parts of the control system and the process control system. After design, the machine is taken into use and the overall safety is validated. During the machine lifetime, there might arise a need for modification or retrofitting the machine with newer technology, which can trigger a return to the relevant lifecycle phase.

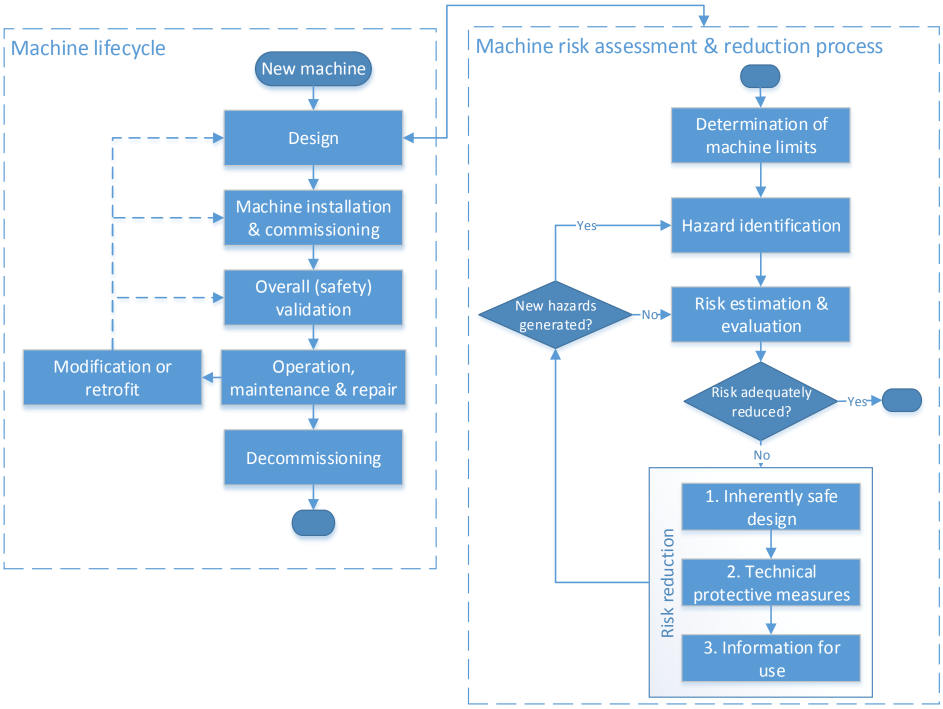

Machine lifecycle
New machine
Design
Machine installation & commissioning
Overall (safety) validation
Modification or retrofit
Operation, maintenance & repair
Decommissioning
Machine risk assessment & reduction process
Determination of machine limits
Hazard identification
Yes
New hazards generated?
No
Risk estimation & evaluation
Risk adequately reduced?
Yes
No
Risk reduction
1. Inherently safe design
2. Technical protective measures
3. Information for use

Figure 1: Machine lifecycle and its risk assessment and reduction process [Own illustration based on ISO 12100 (2010) and IEC 61508 (2010)]

4. SAFE COOPERATION OF MATERIAL HANDLING MACHINERY

Combining the design patterns suggested by Furmans et al. (2010) and the concept of dynamic safety contracts, new application use cases can be considered. A problem point in the flexibility of current systems is especially at the material handling system interfaces, where the payload is transported from one type of material handling equipment to another (handover operation). Some storage equipment, such as storage cranes or automated storage and retrieval systems (AS/RS) could be provided with more flexible interfaces if one allows them to make dynamic safety contracts for the duration of load handover scenarios, enabling machinery to exchange safety-related information. One such example is provided in the following section, where a storage crane uses the laser scanners of an AGV for collision avoidance during a load handover use case.

Further use cases can be imagined with cooperating AGVs. Two or more AGVs could cooperate to transport heavier loads, a scenario where sharing the protection field data from one AGV to another to enable collaborative collision avoidance would be possible through dynamic safety contracts between the AGVs for the duration of the operation. Further use cases could include dynamically integrating different attachments to AGVs depending on the current use case, such as picking arms, like PiRo (Colling et al., 2017), for order picking.

The focus here is only on what Trapp & Schneider (2014) refer to as “Safety certificate at runtime”, meaning that we suggest only use cases where the integration scenario is known at design time, only the exact runtime partner is not. The risk related to the application scenario can be assessed at design time and relevant risk reduction measures can be pre-engineered. The machines check at integration time that all assumptions for the safety certificate are fulfilled. Thus, the safety engineering lifecycle remains overall very similar to the current approach discussed in the previous section.

The new application scenarios introduce higher complexity, especially with regards to the safety-related parts of the control systems. Whereas the current machinery systems are often still analysed using failure modes and effects analysis or with the help of checklists, the likely hazards in such cooperative scenarios are caused by undesired interactions between the machinery, not only by component failures. Therefore we suggest that STPA, with its basis in systems theory, would provide a sound basis for the risk analysis of such scenarios. The STPA analysis can be used to drive especially the definition of the dynamic safety contracts and software safety requirements. In order to fulfil current safety standardisation for machinery, for the final systems the STPA will have to be accompanied with a risk graph estimates for the determined loss scenarios to determine the possibly required performance level of the related risk reduction measures, as well as a probabilistic reliability analysis to determine whether the required performance level was reached.

5. STPA FOR A COOPERATIVE MATERIAL HANDLING APPLICATION

In this section, we discuss our initial approach and experiences in using STPA for a cooperative material handling application: automated payload handover between an automated storage crane and an AGV. For the analysis, the STPA process steps (Leveson & Thomas, 2018) were followed:

1. Define purpose of the analysis
2. Model the control structure
3. Identify unsafe control actions
4. Identify loss scenarios

A system description of the analysed application is provided in section 5.1. The rest of the sections follow the same structure as the basic STPA procedure. The analysis presented here is done only for an initial concept for a future cooperative system. The presented analysis is thus preliminary and incomplete, and we leave further iterations of the analysis with a more detailed concept as future work.

5.1 System description

The analysed system concept is based on the ongoing research project KrasS ("Kransystem zur reproduzierbaren, automatischen und sicheren Stapelung von Gitterboxen" – "A crane system for reproducible, automated and safe stacking of pallet cages") at the Institute for Material Handling and Logistics at Karlsruhe Institute of Technology. The original system concept (Bolender, Oellerich, Braun, Golder, & Furmans, 2018) is shown on the left side in Figure 2. The original system concept is based on an electric overhead bridge crane, which is used for automated storage of pallet cages in a similar manner as automated container handling cranes at container terminals. The storage crane is equipped with a load handling device (LHD) capable of grabbing the pallet cage with the use of corner locks. The storage of the pallet cages is done on an area separated from other operations in the building. The separation could be implemented, for example, by using physical barriers with access monitoring or by safety light curtains. Transporting payloads to or from storage is done through a permanently built handover point and a human-machine interface (HMI). The payload is manually transported to the handover point and a storage operation is requested through the HMI. The payload is automatically picked up by the storage crane, during which the prohibited zone for other operations needs to be extended to include the handover point. When retrieving payloads from storage the process steps are similar, just executed in a different order: HMI-request, automated delivery to handover point, followed by manual pickup.

The main benefit of the manual handover is that the process is quite simple, and can be robustly automated given the fixed handover point. The disadvantages are the need for manual labour, additional fixed infrastructure (handover point) and the large prohibited space when the crane travels to the handover point. The latter could arguably be reduced by limiting the movement of the crane to specific areas when outside the storage stack, as well as with an optimised configuration of the physical barriers or opto-electronic protective devices monitoring the handover point.

The new concept, analysed in this work, is shown on the right side in Figure 2. The concept remains the same when it comes to the storage crane and the storage area. In the new concept the fixed handover point is removed. Instead, the payload is directly transported by the storage crane to or from an AGV for retrieval or storage operation, respectively. The personnel protection during the operation is achieved by sharing the AGV protection field status to the storage crane. If a person or object is detected in the AGV protection field during the operation, both machines stop until the field is clear.

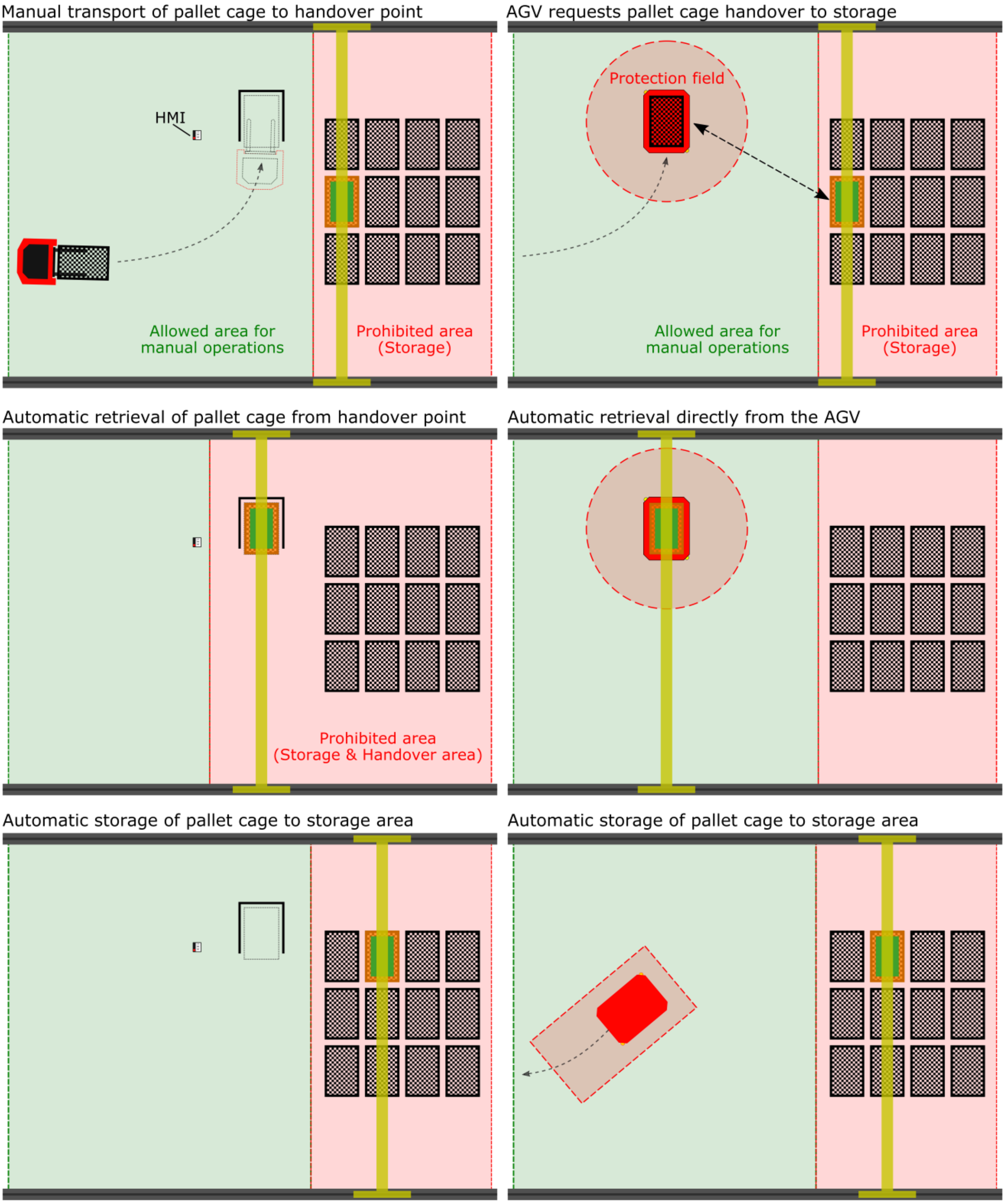

Manual transport of pallet cage to handover point
HMI
Allowed area for manual operations
Prohibited area (Storage)
AGV requests pallet cage handover to storage
Protection field
Allowed area for manual operations
Prohibited area (Storage)
Automatic retrieval of pallet cage from handover point
Prohibited area (Storage & Handover area)
Automatic retrieval directly from the AGV
Automatic storage of pallet cage to storage area
Automatic storage of pallet cage to storage area

Figure 2: Left: Original concept for the pallet cage storage crane with manual handover through a fixed handover point [adopted from (Bolender et al., 2018)]. Right: New concept with storage crane-AGV-cooperation.

The process steps for a handover operation are conceptually as follows:

1) Transport task request and establishing a dynamic safety contract
2) Handover process
    a. For retrieval operation: Storage crane retrieves the payload from storage
    b. Crane drives on top of the AGV
    c. Alignment of AGV for the operation
    d. Handover: LHD is lowered down, grabs or releases the payload and is lifted back up
    e. Once the LHD is back at traveling height, the handover is finished

3) Termination of safety contract, transport task finished, machines return to individual automated operation (For storage operation: Crane stores payload to stack)

The dynamic safety contract here means a short-term safety-related integration of the two machines for the duration of the shared operation. This means opening a real-time communication channel for safety-related communication and sharing of safety-related information. During the integration, both machines will react to protection field breaches detected by the AGV. If an emergency stop-button is activated on either of the machines, they will both react to it. The storage crane is able to issue positioning requests to the AGV to properly align the AGV underneath for the payload lowering or lifting. Additionally the storage crane is able to send a safety-related disable command to the AGV electric drives to prevent AGV movement during the lowering or lifting operation.

For the process description above, some assumptions have been made. There needs to exist a wireless communication channel to enable reliable communication between the two machines. Additionally, there exists a method for the crane to reliably locate the AGV independently of the AGVs self-reported position, this could in practice be achieved for example through a combination of cameras and a real-time locating system. This is required, since the AGVs self-reported position typically cannot be directly used for safety-related functionality. A more detailed technical design of the system is still open; the goal in this work is to utilise STPA on the current concept and use the analysis results for safety-driven design as future work.

The main benefit of the new approach is increased flexibility in material flow operations, with the main disadvantage of increased complexity of the safety automation. The handover can take place anywhere, where enough place for the AGV exists and where the crane can safely drive. The fixed handover place does not need to be installed, saving costs by re-using the existing safety sensors on the AGVs and leaving the surrounding of the crane storage freely re-configurable, depending on the current needs of the production or logistics system supported by the storage crane. A further benefit is, that if the handover interface is standardised, any AGV or, for example, a self-driving lift truck could be used to interface with any storage crane, as long as both were designed capable of the operation.

5.2 Purpose of the analysis

Defining the purpose of the analysis includes identifying unacceptable system-level losses, hazards and constraints. Losses may be anything of value to stakeholders (Leveson & Thomas, 2018). In this work the focus was on the safety-aspects, which are reflected in the defined losses as shown in Table 1. The scope of the analysis is, however, broader than required by the Machinery Directive (European Parliament and The Council of the European Union, 2006) and machinery safety standardisation, as we also consider material losses. The term “machinery” during the analysis refers to the storage crane and the AGV participating in the handover operation.

Table 1 System-level losses

ID	Description
L-1	Loss of life or injury to people
L-2	Loss of or damage to the machinery
L-3	Loss of or damage to the transported goods
L-4	Loss of or damage to other equipment or objects

Prior to defining the system-level hazards, the boundaries of the analysed system need to be identified. In this work we focus on the system consisting of the storage crane and the AGV, which cooperate for the duration of the handover process. An external transport task management system, which could be a central or decentral system, provides transport tasks to the AGVs, which is included in the control diagram, but is not of core interest in the analysis. Additionally there are service/installation personnel, who install and maintain the equipment, and act as a source of the storage area definitions and other parameters input to the storage crane control system.

The system-level hazards are defined in STPA-terminology as system states or conditions, which together with a particular set of worst-case environmental conditions will lead to a loss (Leveson & Thomas, 2018). The ISO 12100 (2010) definition for a hazard is “a potential source of harm”, whereas harm is further defined as “physical injury or damage to health”. The ISO 12100 hazard can further be qualified with respect to its origin, for example, mechanical or electrical hazards. The ISO 12100 is as such more specific than the hazard as understood here. In this analysis we follow the terminology established by Leveson and Thomas (2018) in order to maintain a system-level point of view.

As shown by the system-level hazards in Table 2, the hazards we focus on are mechanical in nature. Typically the most relevant possible hazardous events during operation of material handling machinery are mechanical in nature: moving machinery or payload colliding with people, or the machinery losing control of the payload, which may again result in damage to the payload or it colliding with other objects or personnel. These are the types of hazards, against which the typical safety functions in a material handling machine are built for. Considerations for other types of hazards included in ISO-12100 (2010), such as electrical, noise or ergonomic hazards are outside the scope of the analysis. The system-level constraints derived from the hazards are shown in Table 3.

Table 2 System-level hazards

ID	Description	Link
H-1	Moving parts of the machinery collide with other objects or people during automated operation	L-1, L-2, L-4
H-2	Payload collides with other objects or people during transportation	L-1, L-3, L-4
H-3	Machinery loses control of the payload	L-1, L-2, L-3, L-4

Table 3 System-level constraints

ID	Description	Link
SC-1	Machinery shall maintain separation to other objects and people	H-1
SC-2	Separation of payload to other objects or people during transport shall be maintained	H-2
SC-3	Machinery shall maintain control of the payload during transport	H-3

5.3 Control structure

The control structure used for the analysis is shown in Figure 3. The two main controllers in the system are the storage crane control system and the AGV control system. Each controller has an internal model of the controlled process and a control algorithm, which represents the controller’s decision making process (Leveson & Thomas, 2018). Both of the controllers participate in controlling the material transport process.

The storage crane’s main actuators are the load handling device and the electric drives used to drive the crane axes. The details of the electric drives are abstracted away in the control structure, but we assume that all the axes are driven by frequency converter-controlled motor drives in closed-loop speed control, all equipped with holding brakes to prevent movements when the axes are not actively controlled. The speed feedback used in the electric drives is also communicated to the control system. The operation of the electric drives can also be separately enabled or disabled by the control system. The load handling device is equipped with corner locks, which can be turned open or closed and for each corner lock there is a feedback signal indicating whether the lock is fully turned to locked position or not. For automated operation the crane is equipped with sensors for the position of each axis, the current load mass, as well as access monitoring of the storage area, for example, through a light curtain. As mentioned above, we assume that the crane is able to independently detect the AGV position and alignment underneath. As per safety regulations, both machines are equipped with emergency stop-buttons, which when activated will trigger a safety-related stop of the machine.

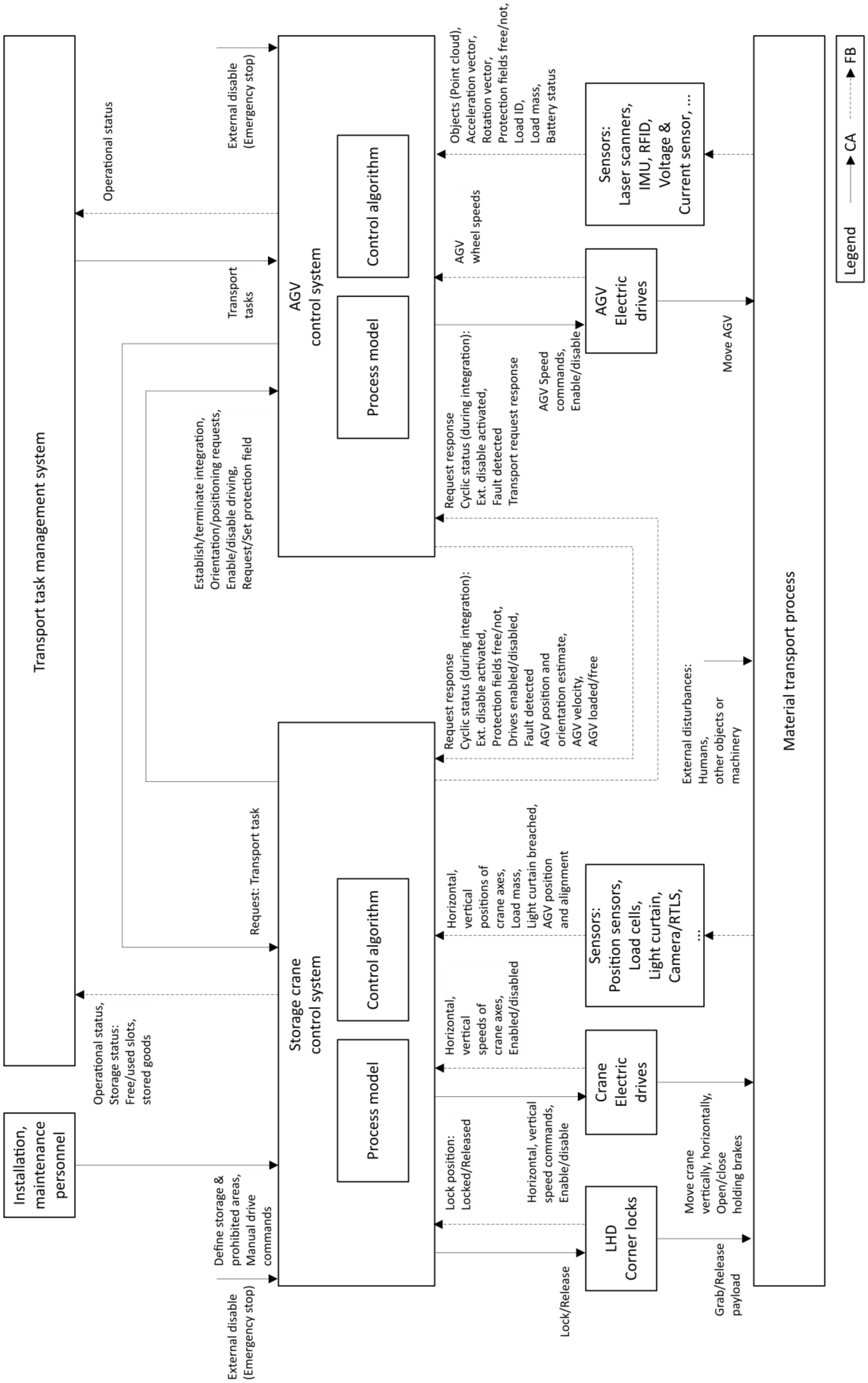

Transport task management system
Installation, maintenance personnel
Operational status
Transport tasks
External disable (Emergency stop)
AGV control system
Process model
Control algorithm
Establish/terminate integration, Orientation/positioning requests, Enable/disable driving, Request/Set protection field
Request: Transport task
Operational status, Storage status: Free/used slots, stored goods
Define storage & prohibited areas, Manual drive commands
External disable (Emergency stop)
Storage crane control system
Process model
Control algorithm
Objects (Point cloud), Acceleration vector, Rotation vector, Protection fields free/not, Load ID, Load mass, Battery status
AGV wheel speeds
Request response Cyclic status (during integration): Ext. disable activated, Fault detected Transport request response
AGV Speed commands, Enable/disable
Request response Cyclic status (during integration): Ext. disable activated, Protection fields free/not, Drives enabled/disabled, Fault detected AGV position and orientation estimate, AGV velocity, AGV loaded/free
Horizontal, vertical positions of crane axes, Load mass, Light curtain breached, AGV position and alignment
Horizontal, vertical speeds of crane axes, Enabled/disabled
Lock position: Locked/Released
Horizontal, vertical speed commands, Enable/disable
Lock/Release
Sensors: Laser scanners, IMU, RFID, Voltage & Current sensor, …
AGV Electric drives
Sensors: Position sensors, Load cells, Light curtain, Camera/RTLS, …
Crane Electric drives
LHD Corner locks
Move AGV
External disturbances: Humans, other objects or machinery
Move crane vertically, horizontally, Open/close holding brakes
Grab/Release payload
Material transport process
Legend
CA
FB

Figure 3: Control structure for storage crane-AGV-cooperation (CA: Control action, FB: Feedback)

The AGV control system is able to drive the AGV with the assistance of its own electric drives, which provide a similar interface as the ones on the crane. The AGV is equipped with laser scanners for collision avoidance, which also provide point cloud data of the environment, which can be used for localisation and mapping with assistance of the data from the inertial measurement unit. The AGV additionally has other sensors for functions such as battery status estimation and load identification. The AGV receives transport tasks from a transport task management system.

To aid in the building of the control structure, as suggested by Leveson and Thomas (2018), responsibilities for the controllers were defined as refinements of the system-level constraints. The responsibilities assigned to the storage crane are listed as an example in Table 4 in Appendix A.

5.4 Unsafe control actions & Loss scenarios

According to Leveson and Thomas (2018), unsafe control actions (UCAs) always consist of five parts: Source, type, control action, context and link to hazards. There are four ways for control actions to be unsafe: 1) Not providing causes a hazard, 2) Providing causes a hazard, 3) Potentially safe control action is provided too early, too late, or in the wrong order, 4) Control actions last too long or are stopped too soon. The second type of UCA can be further considered for contexts in which the action is always unsafe, contexts in which insufficient or excessive actions is unsafe or for contexts in which the direction of the action is unsafe (Leveson & Thomas, 2018).

The control actions from the task management system, maintenance personnel and the external disable signals were not considered in the analysis. Additionally, since the technical details of the protocol for the dynamic safety contract establishment are for now undefined, the details of the contract building were not considered. For the remaining 17 control actions in the control structure, a total of 119 unsafe control actions were identified. An excerpt from the unsafe control actions-table is shown in Table 5 in Appendix B. The UCAs can be used to define controller safety constraints. As future work we will use the analysis results so far to iterate on a more detailed design. The loss scenario-analysis was not done at this point, as it would require accounting for the process models within each controller.

5.5 Discussion

Utilising STPA for the safety analysis of the application scenario proved challenging. Better initial results would have likely been achieved by starting at higher level of abstraction, as the currently included partial details resulted in a large number of identified UCAs. In defining the UCAs, contexts, not strictly related to the cooperative handover application were also included, which further increased the amount of UCAs, making the analysis more complicated than necessary. Regardless of these problems, the executed STPA process steps provided a structured way to identify necessary controls and feedbacks in the concept and the current results seem like a sound basis for further work.

6. CONCLUSIONS AND FUTURE WORK

In this work we suggested using dynamic safety contracts for safe cooperation of machinery to enable new use cases and increased flexibility in material handling operations. We have suggested using STPA for safety-driven design of these integration scenarios and provided initial results of using STPA for a cooperative material handling application. As future work the analysis for the discussed use case will be further iterated towards a detailed technical system concept and design, including definition of the safety functions and contracts required for the cooperation. Additionally, further cooperative scenarios will be studied.

Several open challenges remain. Technical details, such as the specific communication technologies and protocols used are yet to be defined, as well as suitable sensor solutions for various functions, such as for the storage crane to detect the AGV. The STPA analysis results can be used to specify the requirements for the sensors and communication protocols to drive their design. Analysing the system for security-aspects will also be of interest, as security will play an increasingly important role in such applications. As a long-reaching goal one can imagine standardised interfaces for typical integration scenarios to enable flexible cooperative material handling between machines from different manufacturers.

REFERENCES

Abdulkhaleq, A. (2017). *A system-theoretic safety engineering approach for software-intensive systems*: Universität Stuttgart. Retrieved from http://dx.doi.org/10.18419/opus-9049

Bolender, S., Oellerich, J., Braun, M., Golder, M., & Furmans, K. (2018). System zur reproduzierbaren, automatischen und sicheren Stapelung von Gitterboxen mit einem Brückenkran – KrasS. *Logistics Journal: Proceedings*, *2018.* https://doi.org/10.2195/lj_Proc_bolender_de_201811_01

Calinescu, R., Weyns, D., Gerasimou, S., Iftikhar, M. U., Habli, I., & Kelly, T. (2018). Engineering Trustworthy Self-Adaptive Software with Dynamic Assurance Cases. *IEEE Transactions on Software Engineering*, *44*(11), 1039–1069. https://doi.org/10.1109/TSE.2017.2738640

Colling, D., Dziedzitz, J., Hopfgarten, P., Markert, K., Neubehler, K., Eberle, F., . . . Furmans, K. (2017). PiRo - Ein autonomes Kommissioniersystem für inhomogene, chaotische Lager. *Logistics Journal: Proceedings*, *2017.* https://doi.org/10.2195/lj_Proc_colling_de_201710_01

Colling, D., Ibrahimpasic, S., Trenkle, A., & Furmans, K. (2016). Dezentrale Auftragserzeugung und -vergabe für FTF. *Logistics Journal: Proceedings*, *2016.* https://doi.org/10.2195/lj_Proc_colling_de_201610_01

Delfmann, W., Hompel, M. ten, Kersten, W., Schmidt, T., & Stölzle, W. (2018). Logistics as a science: Central research questions in the era of the fourth industrial revolution. *Logistics Research*, *11*(9), 1–13. Retrieved from http://hdl.handle.net/10419/182066

European Parliament and The Council of the European Union (2006). *Directive 2006/42/EC Of The European Parliament And Of The Council of 17 May 2006 on Machinery, and Amending Directive 95/16/EC (Recast)*. (Machinery Directive 2006/42/EC).

Fleming, C. H., & Leveson, N. G. (2014). Improving Hazard Analysis and Certification of Integrated Modular Avionics. *Journal of Aerospace Information Systems*, *11*(6), 397–411. https://doi.org/10.2514/1.I010164

Fleming, C. H., Spencer, M., Thomas, J., Leveson, N., & Wilkinson, C. (2013). Safety assurance in NextGen and complex transportation systems. *Safety Science*, *55*, 173–187. https://doi.org/10.1016/j.ssci.2012.12.005

Furmans, K., Schonung, F., & Gue, K. R. (2010). Plug-and-Work Material Handling Systems. In K. P. Ellis (Ed.), *Progress in material handling research: 2010.* Charlotte: The Material Handling Industry of America. Retrieved from https://digitalcommons.georgiasouthern.edu/pmhr_2010/1/

International Electrotechnical Commission (2010, April 30). *EN IEC 61508 Standard series - Functional safety of electrical/electronic/programmable electronic safety-related systems*. (Standard, EN/IEC 61508:2010).

International Organization for Standardization (2010). *EN ISO 12100:2010 - Safety of machinery - General principles for design - Risk assessment and risk reduction (ISO 12100:2010)*. (Standard, EN ISO 12100:2010).

International Organization for Standardization (2015). *EN ISO 13849-1:2015 - Safety of machinery - Safety-related parts of control systems - Part 1: General principles for design*. (Standard, EN ISO 13849-1:2015).

Leite, F. L., Schneider, D., & Adler, R. (2018). Dynamic Risk Management for Cooperative Autonomous Medical Cyber-Physical Systems. In B. Gallina, A. Skavhaug, E. Schoitsch, & F. Bitsch (Eds.), *Computer Safety, Reliability, and Security* (pp. 126–138). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-99229-7_12

Leveson, N., & Thomas, J. (2018). STPA Handbook. Retrieved from http://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf

Leveson, N. G. (2011). *Engineering a safer world: Systems thinking applied to safety*. *Engineering systems*. Cambridge, Mass.: MIT Press.

Mayer, S. H. (2009). *Development of a completely decentralized control system for modular continuous conveyor systems*. Zugl.: Karlsruhe, Univ., Diss., 2009. *Wissenschaftliche Berichte des Institutes für Fördertechnik und Logistiksysteme der Universität Karlsruhe (TH): Vol. 73*. Karlsruhe: Univ.-Verl. Karlsruhe. Retrieved from https://publikationen.bibliothek.kit.edu/1000011449

Müller, S., & Liggesmeyer, P. (2016). Dynamic Safety Contracts for Functional Cooperation of Automotive Systems. In A. Skavhaug, J. Guiochet, E. Schoitsch, & F. Bitsch (Eds.), *Computer Safety, Reliability, and Security* (pp. 171–182). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-45480-1_14

Popper, J., Blügel, M., Burchardt, H., Horn, S., Merx, J., Richter, D., . . . Staub-Lang, P. (2018). *Safety on modular machines: Whitepaper SF-3.1: 04/2018*. SmartFactoryKL. Retrieved from https://smartfactory.de/wp-content/uploads/2018/04/SF_WhitePaper_Safety_3-1_EN_XS.pdf

Project Consortium KARIS PRO. (2017). *KARIS PRO –Autonomer Materialtransport für flexible Intralogistik: Abschlussbericht des BMBF-Verbundsforschungsprojektes*. Retrieved from http://karispro.de/Abschlussbericht%20KARIS%20PRO.pdf

Rushby, J. (2008). Runtime Certification. In M. Leucker (Ed.), *Runtime Verification: 8th International Workshop, RV 2008, Budapest, Hungary, March 30, 2008. Selected Papers* (pp. 21–35). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-89247-2_2

Rushby, J. (2016). Trustworthy Self-Integrating Systems. In N. Bjørner, S. Prasad, & L. Parida (Eds.): *Lecture Notes in Computer Science, 12th International Conference on Distributed Computing and Internet Technology, ICDCIT 2016* (pp. 19–29). Bhubaneswar, India: Springer-Verlag. Retrieved from https://doi.org/10.1007/978-3-319-28034-9_3

Schneider, D. (2014). *Conditional safety certification for open adaptive systems*. *PhD Theses in Experimental Software Engineering: Vol. 48*. Stuttgart: Fraunhofer-Verl. Retrieved from http://publica.fraunhofer.de/documents/N-283653.html

Seibold, Z. (2016). *Logical Time for Decentralized Control of Material Handling Systems*. *Wissenschaftliche Berichte des Instituts für Fördertechnik und Logistiksysteme des Karlsruher Instituts für Technologie (KIT): Vol. 89*: KIT Scientific Publishing. Retrieved from https://publikationen.bibliothek.kit.edu/1000057838

Trapp, M., & Schneider, D. (2014). Safety Assurance of Open Adaptive Systems - A Survey. In N. Bencomo, R. France, B. H. C. Cheng, & U. Aßmann (Eds.), *Models@run.time: Foundations, Applications, and Roadmaps* (pp. 279–318). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-08915-7_11

Ullrich, G., & Kachur, P. A. (2015). *Automated guided vehicle systems: A primer with practical applications* (Second revised and expanded edition). Berlin: Springer.

APPENDIX A. CONTROLLER RESPONSIBILITIES

Table 4 Storage crane control system: Responsibilities and links to system-level constraints

ID	Responsibility Description	Link
R-1.1	LHD/Payload lifting or lowering shall only be allowed above the pre-defined storage area or above an AGV during a handover	SC-1, SC-2
R-1.2	Maintain LHD/Payload at top-most lifting position during horizontal movements	SC-1-3
R-1.3	Prevent lifting if detected load mass exceeds 110% of maximum load	SC-3
R-1.4	Maintain velocity below maximum allowed velocity for all axes	SC-3
R-1.5	Maintain crane position between the allowed limits for all axes	SC-1-3
R-1.6	Stop and prevent crane movements if light curtain to storage area breached	SC-1, SC-2
R-1.7	Stop and prevent crane movements if external disable is active	SC-1, SC-2
R-1.8	Monitor internal sensors and actuators, stop and prevent crane movement if internal	SC-1-3

	fault has been detected	
R-1.9	All corner locks shall be locked before lifting the payload	SC-3
R-1.10	Corner locks shall not be opened before the payload is lowered on the storage area or on an AGV	SC-3
R-1.11	Enforce maximum stacking height as allowed by the payload	SC-3
R-1.12	During handover, stop and prevent crane movements if AGV protection field is breached	SC-1, SC-2
R-1.13	During handover, stop and prevent crane movements if the AGV moves unrequested	SC-1-3
R-1.14	During handover, stop and prevent crane movements if external disable is activated through the AGV	SC-1, SC-2
R-1.15	During handover, stop and prevent crane movements if fault detected is reported by the AGV	SC-1-3
R-1.16	During handover, issue orientation/positioning requests to the AGV until proper alignment under the crane has been detected	SC-3
R-1.17	During handover, monitor the timing of cyclic status updates from the AGV, stop and prevent crane movement if the timing requirements (TBD) are not fulfilled	SC-1, SC-2

APPENDIX B. EXCERPT FROM UNSAFE CONTROL ACTION TABLES

Table 5 Excerpt from UCA-tables (SCCS: Storage crane control system, LHD CL: Load handling device corner locks, AGVCS: AGV control system, AGVED: AGV electric drives)

CA	Not providing causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, Applied too long
CA-1: Lock/release (SCCS to LHD CL)	UCA-1.1: SCCS does not provide lock command to all corner locks when attempting the grab a payload [H-2, H-3] UCA-1.2: SCCS does not provide release command to all corner locks when attempting to release a payload [H-2, H-3]	UCA-1.3: SCCS provides release command while payload is grabbed and lifted in air [H-2, H-3]	UCA-1.4: SCCS provides release commands while payload has not been fully lowered [H-2, H-3] UCA-1.5: SCCS provides lock-commands while LHD has not yet been fully lowered on the payload [H-2, H-3] UCA-1.6: SCCS provides lock/release commands while upward driving command has already been issued (LHD/Payload still fully/partially on ground) [H-2, H-3]	UCA-1.7: SCCS stops providing lock-command to CLs before all are locked during payload grabbing [H-2, H-3] UCA-1.8: SCCS stops providing release-command to CLs before all are released during payload releasing [H-2, H-3]
CA-10: Enable/Disable drives (AGVCS to AGVED)	UCA-10.1: AGVCS does not disable driving while external disable (directly or through crane during handover) is active [H-1, H-2] UCA-10.2: AGVCS		UCA-10.6: AGVCS disables driving during handover while alignment-phase is not yet finished. [H-1, H-2, H-3] UCA-10.7: AGVCS disables driving too	UCA-10.8: AGVCS re-enables driving before handover operation is finished (crane still relies on protection field) [H-1, H-2]

	does not disable driving while protection fields are breached [H-1, H-2] UCA-10.3: AGVCS does not disable driving when maximum allowed velocity has been exceeded. [H-1, H-2, H-3] UCA-10.4: AGVCS does not disable driving while internal faults, communication faults detected or fault reported by crane during handover [H-1, H-2, H-3] UCA-10.5: AGVCS does not disable driving when requested by the crane during the handover operation [H-1, H-2, H-3]		late after commanded by SCCS during handover [H-1, H-2, H-3]	

A comparative safety assessment for Direct Current and Direct Current with hybrid supply power systems in a windfarm Service Operation Vessel using System-Theoretic Process Analysis

Victor Bolbot[1*], Romanas Puisa[1], Gerasimos Theotokatos[1], Evangelos Boulougouris[1] and Dracos Vassalos[1]
[1] Maritime Safety Research Centre, University of Strathclyde, UK

ABSTRACT

As windfarms are moving further offshore, their maintenance has to be supported by the new generation Service Operation Vessels (SOV) with Dynamic Positioning capabilities. For the SOV safe operations it is crucial that any hazardous scenario is properly controlled. Whilst international regulations require the implementation of Failure Modes and Effects Analysis (FMEA) for SOV power systems, FMEA has been criticised for not addressing properly failures in control systems. In this study, System-Theoretic Process Analysis (STPA) is employed for identifying the hazardous scenarios in terms of Unsafe Control Actions (UCAs) in Direct Current (DC) and DC with batteries power systems. Then the identified UCAs are ranked based on their risk. The results demonstrate that the number of hazardous scenarios derived by the STPA increases in a power system with batteries in comparison to a conventional DC power system, thus depicting higher complexity of this system. However, the increase in overall risk is small and within acceptable limits, whilst the risk reduces for a number of UCAs leading to Diesel Generator overload sub-hazard.

Keywords: Windfarm Service Operation Vessels, Safety, Blackouts, Diesel-Electric Propulsion, Hybrid Diesel-Electric Propulsion

1 INTRODUCTION

Offshore wind-faming is becoming a major source of the renewable energy in many countries. However, the offshore wind farms maintenance cost currently impacts on the competitiveness of the electricity produced. Present safety requirements and needs of the service personnel influence wind farm locations and operational flexibility. Consequently, future Service Operation Vessels (SOVs) need to be more efficient and safer in order to meet future demands. Next generation support vessels providing safe and more efficient offshore wind farm servicing (the EU-funded NEXUS project) is aiming to deliver an advanced SOV design optimised for efficiency, performance, safety, and working environment whilst minimising costs throughout the life-cycle by 20% compared to the current state of the art vessels (EC, 2019). As wind farms are moving further from the coast, significant innovations in the SOV design are required. This, together with stringer emission regulations and fluidity in the fuel market prices, render attractive the use of alternative fuels and power generation systems, including hybrid power supply, where diesel-generators and batteries are used to cover ship energy needs.

The incorporation of batteries achieves fuel consumption reduction by running Diesel Generator (D/G) sets at optimum load by peak load shaving and functioning as a spinning reserve

---

[*] Corresponding author: tel. +447706578021 email: victor.bolbot@strath.ac.uk

(Brandsaeter, Valoen, Mollestad, & Haugom, 2015; Geertsma, Negenborn, Visser, & Hopman, 2017; Räsänen, 2017). Implementation of batteries support the D/G sets downsizing, which results in the D/G sets operation at their most efficient load ranges (Brandsaeter et al., 2015). Other advantages include higher redundancy in the system and lower emissions due to the batteries charging from the local grid in harbour (Brandsaeter et al., 2015; Geertsma et al., 2017). On the SOV, due to the Dynamic Positioning (DP) power requirements, the D/G sets are often oversized or pushed to operate at lower loads to be able to withstand a sudden loss of a D/G set in adverse weather conditions. Therefore, incorporation of batteries to provide the necessary spinning reserve during faulty conditions or power during power peaks on SOV can provide substantial benefits in terms of fuel savings during DP and other operations. Batteries disadvantages include relatively high procurement cost (Brandsaeter et al., 2015; Geertsma et al., 2017), large batteries size and weight (Räsänen, 2017), limited number of recharging cycles (Räsänen, 2017) and addition of new hazardous scenarios to the system (Bolbot, Theotokatos, Boulougouris, & Vassalos, 2019; Brandsaeter et al., 2015).

On the next generation SOV, with increased technicians and crew numbers, ensuring safety of power generation system is paramount as any malfunctions such as blackout or brownout may lead to contact/collision/grounding. These accidents in turn can result in ships progressive flooding and capsize with crew and technicians getting drown (Vassalos et al., 2019). In addition, the introduction of batteries increases hazardous scenarios number resulting in fire, explosion and crew intoxication (Brandsaeter et al., 2015), e.g., a fire on hybrid-electric tugboat occurred due to malfunction of Battery Management System (Hill, Agarwal, & Gully, 2015), whilst a number of similar incidences have been reported in other industries (Hill et al., 2015). In this respect, it is crucial to ensure that all these scenarios are identified and properly addressed during the system design.

The primary reference for designing safe power generation systems is the IMO regulations (Organization, 2014) and classification society rules (DNVGL, 2015). Currently, the main hazard identification method in the DP systems is the Failure Mode and Effect Analysis (FMEA), which is applied to ensure adequate system components redundancy (DNVGL, 2015; IMCA, 2015). In previous studies, a high-level FMEA has been used for comparative safety analysis of different propulsion systems, including power system with batteries in other ships, for example a Ferry boat in (Jeong, Oguz, Wang, & Zhou, 2018). However, FMEA has been criticised for not addressing properly the automation functions in the system (Bolbot, Theotokatos, Bujorianu, Boulougouris, & Vassalos, 2019; Rokseth, Utne, & Vinnem, 2017; Sulaman, Beer, Felderer, & Höst, 2017; Thomas, 2013). On the other hand, control and automation functions have an important role for power generation on DP vessels (United Kingdom Protection & Indemnity Club, 2015). Considering this, System-Theoretic Process Analysis (STPA) has been proposed to be used to address the complexity in interactions between the control systems and physical processes (N. G. Leveson, 2011). In (Bolbot, Theotokatos, Boulougouris, et al., 2019) the safety of hybrid-electric propulsion system and classical propulsion system using Alternate Current for electrical power distribution has been compared using STPA on a cruise ship vessel. Other studies have referred to potential safety issues on ship power systems with batteries but they did not follow a hazard identification method for their analysis (Hill et al., 2015).

Pertinent literature reveals a number of research gaps: (a) hazard analysis of power systems with Direct Current (DC) power network and DC power with batteries system on SOV using STPA and (b) incorporation of risk as a measure in STPA to compare different designs. The research gap leads to the aim of this study, which is to analyse the safety of power systems on SOV with batteries using STPA and to compare it with standard DC power systems in terms of risk.

This paper is organised as follows: in section two, the methodology steps are presented; in section three, a short description of the analysed system is provided; in section four, the analysis results and safety recommendations are given; finally, in section five, the main findings of this study are summarised.

2 METHODOLOGY

As it has been referred in the introduction, STPA has been selected in this study to identify the hazardous scenarios. However STPA has been criticised for not allowing risk estimation and criticality analysis (Dawson et al., 2015); for this reason the STPA method has been enhanced. The method steps are presented in Figure 1 and described in more detail below.

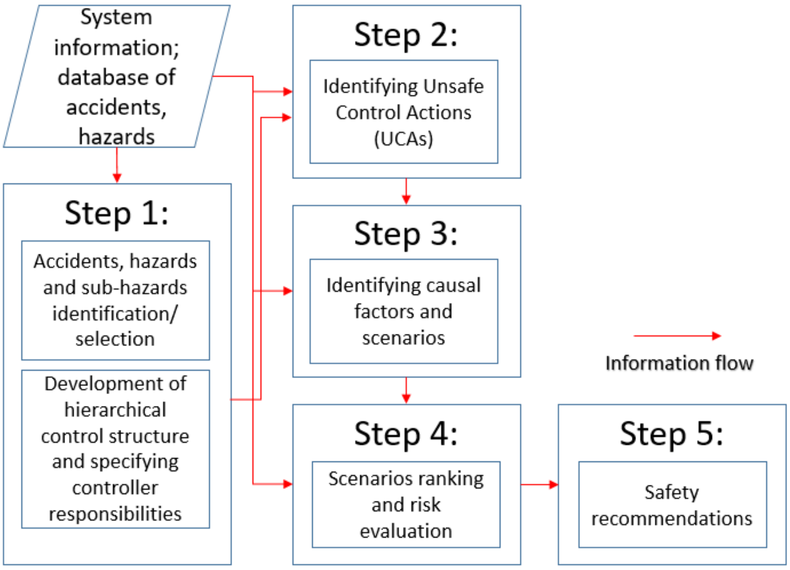

System information; database of accidents, hazards
Step 1:
Accidents, hazards and sub-hazards identification/ selection
Development of hierarchical control structure and specifying controller responsibilities
Step 2:
Identifying Unsafe Control Actions (UCAs)
Step 3:
Identifying causal factors and scenarios
Step 4:
Scenarios ranking and risk evaluation
Step 5:
Safety recommendations
Information flow

Figure 1 STPA steps.

STPA defines the accident as: "an undesired and unplanned event that results in loss, including loss of human life or human injury, property damage, environmental pollution, mission loss, financial loss, etc." (Leveson & Thomas, 2018). The hazards in the STPA framework are understood as: "system states or set of conditions that together with a worst-case set of environmental conditions, will lead to an accident" (Leveson & Thomas, 2018). The hazards in STPA are viewed on a system level, so they go beyond the single failures that may occur in the system and should be referred to a specific state of the system. Sub-hazards are considered states in a worst-case scenario leading to hazard realisation. Generic requirements can be specified, based on the hazards and sub hazards.

The development of a functional control structure is one of the differentiating points of the STPA analysis, compared with the other methods (Leveson & Thomas, 2018). Usually, it starts with a high-level abstraction of the system and proceeds to a more detailed system description. The initial control structure consists of the high-level controller, the human operator and the controlled process with the basic control, feedback and communication links. A more detailed description would incorporate a hierarchy of controllers. Both high-level and detailed control structure can be used for the safety analysis at different system design stages. After the development of the basic control structure, the next step is its refinement. The required actions include a) the identification of each controller responsibilities; b) the process model with process variables and potential process variable values; c) the control actions; d) the behaviour of the actuators; e) the information from the sensors; f) the information from the other controllers.

The actual hazards identification starts by finding the Unsafe Control Actions (UCAs). The possible ways to proceed are either by using the control actions types as initially proposed for the STPA (N. Leveson, 2011) or by using the context tables as proposed in Thomas (2013). Herein, the second of the two approaches has been selected. According to both approaches, the possible UCAs can be of the following seven types (Leveson & Thomas, 2018):

- Not providing the action leads to a hazard.
- Providing of a UCA that leads to a hazard.
- Providing the control action too late.

- Providing the control action too early.
- Providing the control action out of sequence.
- Control action is stopped too soon
- Control action is applied for too long.

According to the STPA, there is also another type of UCA, when the safe control action is provided but is not followed. This type of failure mode is addressed during the identification of causal factors in the second step of the method. Similarly, with the system hazards, safety constraints can be derived for the UCAs, aiding the identification of possible safety barriers.

The second step in the hazard identification of the STPA has the purpose of determining all the scenarios and causal factors leading to the UCAs. This is done by examining the hazardous scenarios, including software and physical failures as well as design errors. There are several ways to organise the results of the hazardous scenarios by using tables or lists. In this work, the process was augmented by a checklist, developed on the basis of previous studies (Becker & Van Eikema Hommes, 2014; Blandine, 2013). The main categories of causal factors are:

- Inappropriate control input
- Hardware failure
- Software faulty implementation
- Software faulty design
- Erroneous or missing input
- Inadequate control command transmission
- Flawed execution due to failures in actuator or physical process
- Conflicting control actions

The systemic and contributory causal factors (Puisa, Lin, Bolbot, & Vassalos, 2018) have not been considered during identification of the causal factors, as the implementations of proper training for system operator and maintenance is out of the scope of system designer. The aim of the designer is to ensure the adequate reliability and availability of system functions. Therefore, the aim of the analysis is to rank the different hazardous scenarios identified by the STPA to allow better allocation of resources to specific controllers; hence the different scenarios (UCAs) risk is estimated.

The new part of the STPA in the presented methodology is the risk estimation for the identified UCAs. The basic assumption behind the estimation is that UCA can be considered as the central undesired event in the system, thus being in the centre of the Bow Tie as depicted in Figure 2. Then the total risk can be estimated as aggregation of individual UCAs risks. In a similar way with Level of Protection Analysis method (BSI, 2004), the risk of an UCA is considered dependent on its causal factors, the effectiveness of mitigation barriers, and coincidence with inadvertent environmental factors. If the causal factors likelihood, the accident severity, the mitigation barriers/measures effectiveness and relevant inadvertent environmental factors are quantified, the risk for each UCA can be estimated.

For the analysis presented in the methodology herein, with the exception of the above, the following additional assumptions have been made:

- The UCAs causal factors are independent (Blandine, 2013) as the systemic and contributory factors (Puisa et al., 2018) are omitted as the focus is on the system design.
- If UCA leads to more than two hazards, then paths with the smaller risk can be ignored.
- Similarly, if multiple causal factors result in UCAs, the causal factors with smaller likelihood can be ignored for estimations.
- The overall risk can be aggregated and calculated for the system based on individual UCAs risk.
- Each mitigation barrier can mitigate the 90% of relevant hazardous conditions. This is rather a conservative assumption with regard to effectiveness of mitigation barriers (BSI, 2004).
- The UCA causal factors frequency and the UCA context factors frequency are independent from each other.

- The UCA causal factors frequency is estimated by considering it together with the relevant UCAs preventative barriers effectiveness.
- Accidents are considered as disjoint and independent.
- If UCAs are caused by other UCAs (they are practically their causal factors), then these causal UCAs are omitted for estimation of risk for UCAs. Instead, these causal UCAs are considered to contribute to risk independently from other UCAs.
- Causal factors resulting in multiple UCAs occurring are repeated for each UCA risk estimation, as this assumption has no influence on estimation of the total risk.

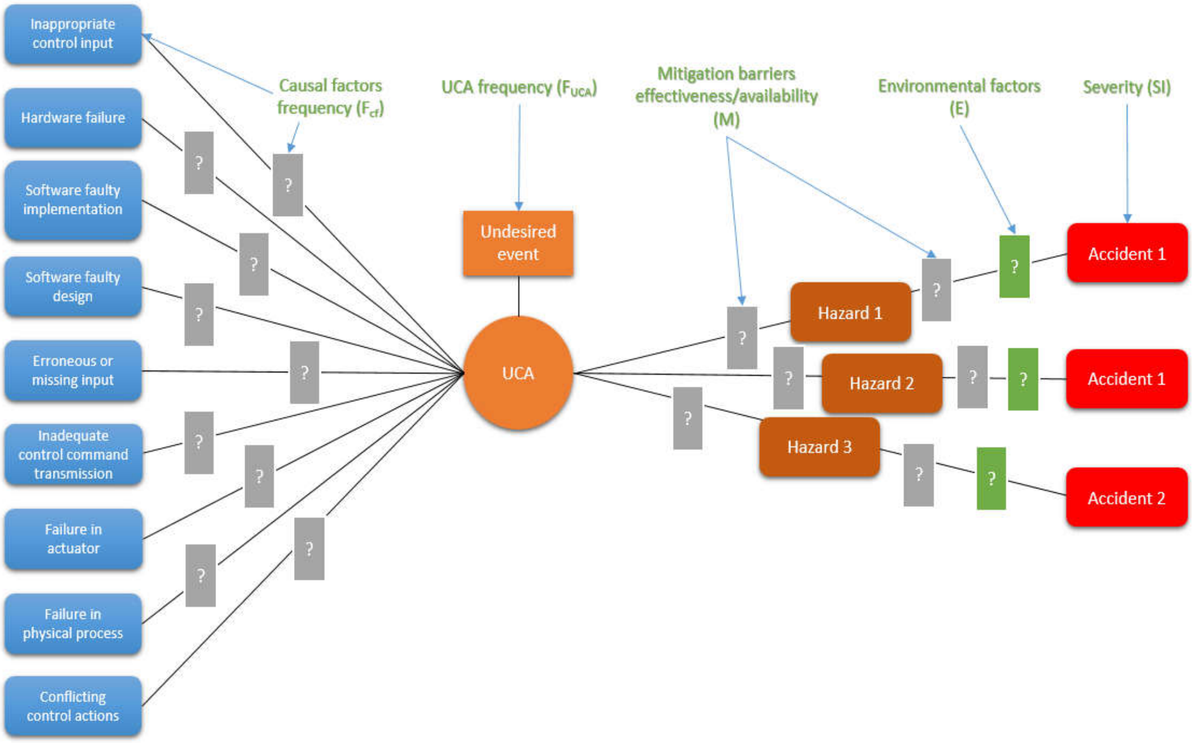

Inappropriate control input
Hardware failure
Software faulty implementation
Software faulty design
Erroneous or missing input
Inadequate control command transmission
Failure in actuator
Failure in physical process
Conflicting control actions
Causal factors frequency (F_{cf})
UCA frequency (F_{UCA})
Mitigation barriers effectiveness/availability (M)
Environmental factors (E)
Severity (SI)
Undesired event
UCA
Hazard 1
Hazard 2
Hazard 3
Accident 1
Accident 1
Accident 2

Figure 2 The simplified Bow Tie

The Potential Loss of Life (PLL) is one of the expressions of Societal Risk (International Maritime Organisation, 2013) and is defined as expected value of the number of fatalities per year (International Maritime Organisation, 2013; Vinnem, 2014):

$$PLL = \sum_l \sum_j f_{lj} c_{lj} \tag{1}$$

Where f_{lj} is the annual frequency of accidental scenario (event tree terminal event) l with personal consequences j and c_{lj} is expected number of fatalities in each accidental scenario (event tree terminal event) l with personal consequences j.

The PLL is connected to the Individual Risk (IR) according to the following equation (Johansen & Rausand, 2012), where N is the number of people in population exposed to risk:

$$PLL = N\ IR \tag{2}$$

Based on the assumptions above, the PLL can be approximated as sum of risk of n individual UCAs as follows:

$$PLL_{app} = \sum_k^n R_k \tag{3}$$

Now the risk R_k for each UCA using f_k frequency of accidental scenario and c_k consequence of accidental scenario expressed in fatalities per year is estimated as follows:

$$R_k = f_k \times c_k \qquad \text{[fatalities per ship-year]} \qquad (4)$$

The frequency of each accidental scenario is estimated using UCA frequency F_{UCA}, effectiveness of mitigation controls M and probability of inadvertent environmental context E as in eq. (5) and the severity of each accidental scenario is estimated as in eq.(6):

$$f_k = F_{UCA} \times M \times E = F_{UCA} \times 10^{M_k-6} \times 10^{E_k-3} \qquad \text{[events per ship-year]} \, (5)$$
$$c_k = 10^{SI_k-3} \qquad \text{[fatalities per events]} \quad (6)$$

The ranking M_k for effectiveness of mitigation measures is implemented according to Table 1. For the ranking M_k of available mitigating barriers, different mitigating barriers type are considered namely a) the presence of redundant component implementing the same function with the faulty one, b) available safety or reconfiguration functions c) humans operators rectification actions. The ranking of inadvertent environmental factors (E_k) is implemented as in Table 3. The Severity Index for accident (SI_k) is selected according to Table 2 retrieved from Formal Safety Assessment Guidelines (International Maritime Organisation, 2013).

The UCA is described by referring to the controller, the control action, the control action failure type, the context and the link to the hazard (Leveson & Thomas, 2018). Practically though, an UCA will occur if specific control action failure mode is realised in specific context. In case of a Fault Tree this relationship would be represented using AND gate, hence multiplication between frequency of control action failure mode and probability of specific context is required. However, the control action failure mode can be attributed to the specific causal factors, identified previously, which can be connected using OR gate to the UCA (Blandine, 2013). Wrong execution practically refers to one of the UCAs types (Leveson & Thomas, 2018) and has been already included in identification of causal factors. Therefore, the UCAs frequency (F_{UCA}) is estimated as in eq.(7) using frequency of causal factors F_{cf} leading to relevant control action failure mode, the number of controllers m in system, which can implement the specific UCA and the probability of the UCA context:

$$F_{UCA} = m \times Max(F_{cf}) \times 10^{UC_k-4} \qquad \text{[events per ship-year]} \, (7)$$

The F_{cf} is ranked using Table 4, retrieved from Formal Safety Assessment Guidelines (International Maritime Organisation, 2013) and is estimated as in eq.(8), whilst UC_k ranking used for estimating the probability of UCA context is based on Table 5.

$$F_{cf} = 10^{FI_{kj}-6} \qquad \text{[events per ship-year]} \, (8)$$

Table 1 Ranking for availability of UCAs mitigation measures

Ranking (M_k)	Definition	Unavailability of mitigation measures
6	No controls provided	10^{-0}
5	Some mitigation controls availability (One control barrier)	10^{-1}
4	Adequate mitigation controls availability (Two control barriers)	10^{-2}
3	Rare mitigation controls unavailability (Three control barriers)	10^{-3}
2	Remote mitigations controls unavailability (Four control barriers)	10^{-4}
1	Extremely remote mitigations controls unavailability (Five control barriers and above)	10^{-5}

Table 2 Ranking for severity of UCAs hazards/accidents (International Maritime Organisation, 2013).

Ranking (SI_k)	Definition	Effects on human Safety	Effects on ship	Oil spillage	Equivalent fatalities
4	Catastrophic	Multiple fatalities	Total loss	Oil spill size between < 100 - 1000 tonnes	10
3	Severe	Single fatality or multiple severe injuries	Severe damage	Oil spill size between < 10 - 100 tonnes	10^{-0}
2	Significant	Multiple or severe injuries	Non-severe ship damage	Oil spill size between < 1 - 10 tonnes	10^{-1}
1	Minor	Single or minor injuries	Local equipment damage	Oil spill size < 1 tonne	10^{-2}

Table 3 Ranking for inadvertent environmental factors.

Ranking (E_k)	Definition	Probability of inadvertent environmental factors
3	Uncontrolled UCA will always lead to accident	10^{-0}
2	Uncontrolled UCA will sometimes lead to accident	10^{-1}
1	Uncontrolled UCA will rarely lead to accident	10^{-2}

Table 4 Ranking for causal factors frequency (International Maritime Organisation, 2013).

Ranking (FI_{kj})	Definition	F (per ship year)	F (per ship hour)
7	Likely to occur once per month on one ship	10	1.14 10^{-3}
5	Likely to occur once per year in a fleet of 10 ships, i.e. likely to occur a few times during the ship's life	10^{-1}	1.14 10^{-5}
3	Likely to occur once per year in a fleet of 1,000 ships, i.e. likely to occur in the total life of several similar ships	10^{-3}	1.14 10^{-7}
1	Likely to occur once in the lifetime (20 years) of a world fleet of 5,000 ships	10^{-5}	1.14 10^{-9}

Table 5 Probability of UCA context.

Ranking (UC_k)	Definition	Probability of inadvertent environmental factors
4	Always	10^{-0}
3	Sometimes	10^{-1}
2	Rarely	10^{-2}
1	Remotely	10^{-3}

3 CASE STUDY DESCRIPTION

The initial power system and hybrid-electric power system single line diagram are presented in Figure 4 whilst the functional control structure for both systems is given in Figure 3. Two switchboards and engine rooms are required to comply with the DP requirements. The power network is of the Direct Current type. Power Management System (PMS) starts/stops the engines based on the ship consumers electric load demand. Switchover between the plant Diesel Generators (D/G) is implemented based on the D/G sets running hours. The PMS can implement a fast-electrical load reduction for the propulsion motors and bow thrusters as well as preferential tripping functions (fast load reduction) by tripping electrical consumers. The D/G sets can operate in the variable speed mode and their power output is regulated by speed governor (ECU 7) and Automatic Voltage Regulator (AVR) whilst delivered power to network through converters is controlled by the Generator Control Unit (GCU). A number of other smaller functions are supported by EIM and EMU units on the D/G sets. Power transferred between sections is controlled by Bus Tie Unit (BTU). Several safety systems are used to trip the D/G sets and the propulsion motors if a fault had been observed.

In the investigated hybrid-electric power system, in addition to the initial system components, one battery pack per switchboard is installed. The battery output and condition are controlled by a dedicated Battery Management System (BMS), which monitors the actual battery health state, the battery and cell capacity and controls the battery cells charge status, the discharging/charging rate, the power output and the battery auxiliary systems. The BMS communicates with PMS to determine the actual power status and power demand implementing in this way the Energy Management System functions. The BMS also communicates with fire-fighting systems to ensure the firefighting

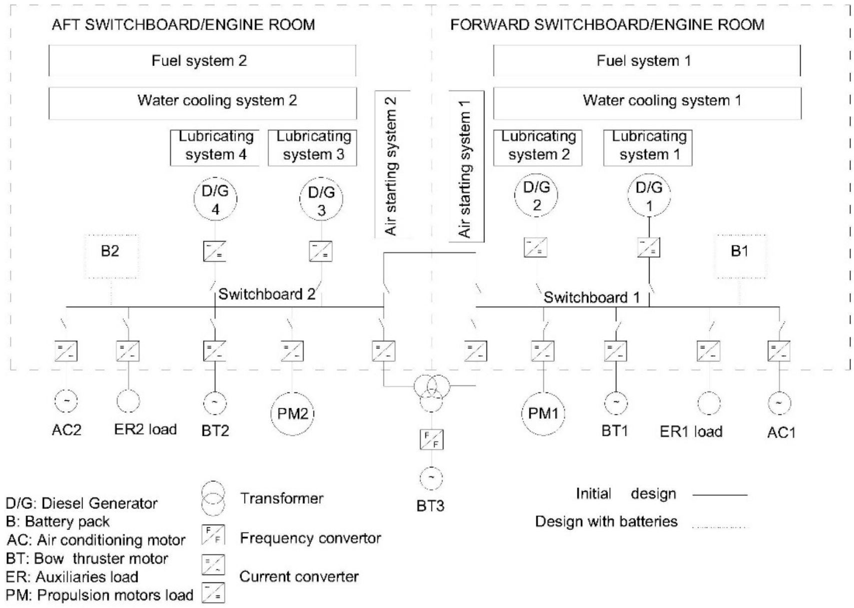


Figure 4 Power network layout diagram

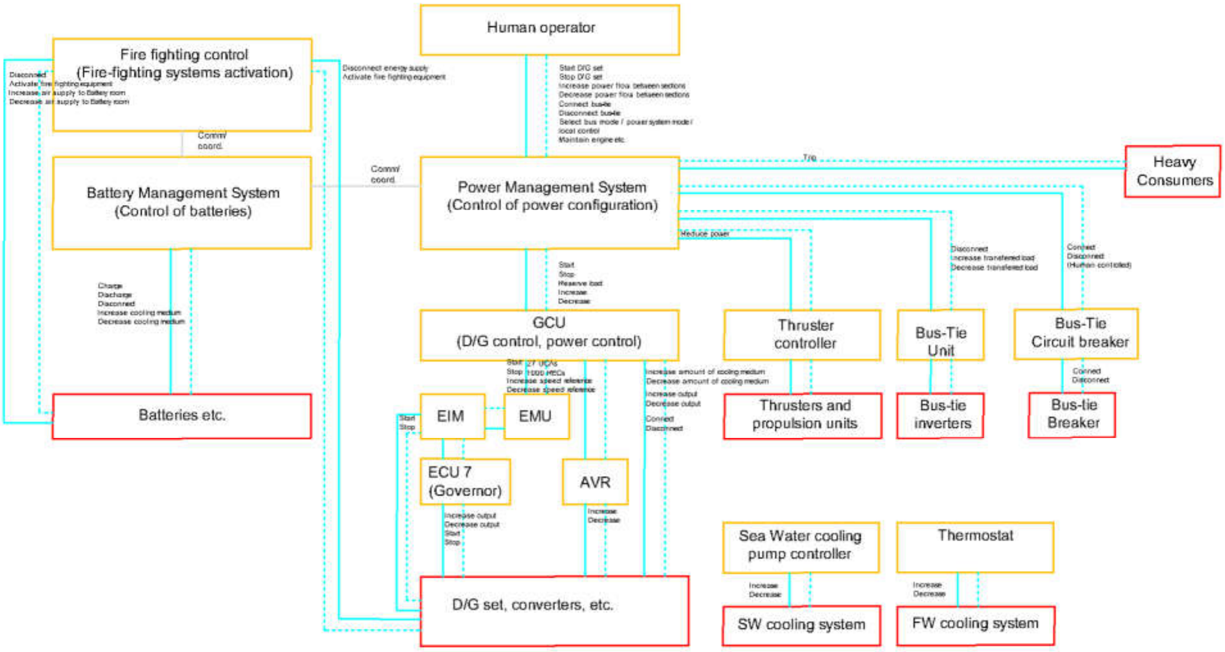


Figure 3 Power network control structure

actions operation. Battery capacity is considered adequate to cover the whole ship power demand for a limited period. The considered battery is of Li-Ion type.

The following has been assummed with respect to the systems operation:

- The power system control network is isolated from other networks, so no hazardous scenarios are developed in the system because of cyber-attacks.
- The human operator does not introduce new hazards, only mitigates them.
- Power plant operates with the bus-tie circuit breaker disconnected.
- Power can be transferred from switchboard to a switchboard using converters at Bow thruster motor 3.

With respect to the case study it has assumed that the SI_k for each UCA is either 2 (Significant) or 3 (Severe). In addition the number of people on the ship, including crew and technicians has been estimated as 60.

4 RESULTS AND DISCUSSION

Based on previous Formal Safety Assessment studies, the following causality scenarios can be considered as accidents (IMO, 2008):

- Collision [A-1]
- Contact [A-2]
- Grounding [A-3]
- Fire [A-4]
- Explosion [A-5]
- Machinery damage [A-6]
- Foundering [A-7]
- Operating personnel injury or death [A-8]

These accidents are not fully disjoint, as a fire can lead to collision and vice versa (Hamann, Papanikolaou, Eliopoulou, & Golyshev, 2013). In addition, numerous hazards can be connected to the accidents on a cruise ship and there can be interactions between different hazards. Herein, the most important and those related to the system under analysis are referred (Bolbot, Theotokatos, & Vassalos, 2018; IMO, 2008):

- Propulsion loss [H-1] leading to collision, contact and grounding accidents. The propulsion loss can be further developed into the following sub-hazards:
    - D/G sets overload [H-1-1].
    - Transients [H-1-2].
    - Imbalanced power generation [H-1-3]
    - D/G sets unavailability [H-1-4]
    - Batteries unavailability [H-1-5]
    - Propulsion motors unavailability [H-1-6]
- Conditions contributing to fire in the engine room [H-2].
- Uncontrolled electrical faults in equipment leading to [H-3] fire and explosions in system components or blackout (propulsion loss).
- Toxic/flammable atmosphere in battery room leading to crew intoxication and/or fire [H-4].
- Anomalous conditions in batteries leading to fire and its expansion [H-5].
- Arson – deliberate act resulting in fire [H-6].
- Human erroneous operation [H-7]
- Cyber-attack leading to any of previous hazards [H-8].
- Water ingress [H-9]

Although, it is acknowledged that there is contribution from hazards [H-6]-[H-9] to the overall system risk, these hazards can be considered as external to the system presented in Figure 4 and Figure 3 and thus their analysis has been omitted. The interconnection between hazards and accidents is schematically shown in Figure 5.

The developed control structure has been already provided in Figure 3. The difference between the two power systems can be found in the presence of Battery Management System and additional interactions between the fire-fighting system and the power system. The description of responsibilities of each controller and their control actions, although necessary and used for the analysis, have been omitted for brevity and confidentiality purposes.

The results of applying STPA and risk analysis and comparing the different results are presented in Table 7, Table 8, Figure 6 and Figure 7. A guiding example of application of the method is provided in Table 6. As it can be observed from Table 7, the number of the UCAs and the associated causal factors is significantly higher in the system with batteries. This is owed to the increased number of interactions between the control systems and the physical processes in a power system with batteries. However, the estimated risk is only slightly higher in the power system with batteries. The estimated individual risk for different Severity Indexes is smaller than negligible 10^{-6} and in every case smaller than the maximum tolerable risk for the crew 10^{-3} and maximum tolerable risk for passengers 10^{-4} (International Maritime Organisation, 2013). So it can be considered as acceptable. However, it should be noted that the estimated risk includes only failures in control systems, whilst some scenarios that could be potentially identified with FMEA have not been addressed. Consequently, the estimated risk would be greater, if FMEA related accidental scenarios have been incorporated. It should be also noted, that there is a specific subjectivity in the analysis, as a) uncertainty in the estimated frequencies and probabilities has not been incorporated and b) there are numerical approximations in calculations due to the use of tables with rankings. Consequently, the estimated risk must be taken with precaution. The subjectivity that exists in the risk assessments is one of its major weaknesses (Aven, 2016; Goerlandt, Khakzad, & Reniers, 2016). Last, but not least the risk is estimated for a system and not the whole vessel, so it can be used for comparison with acceptable values with precaution; it can be used though for comparison of different systems and scenarios.

As it can be observed from the Table 8, the incorporation of batteries reduces the risk in all the controllers but BMS. In addition, from the Figure 6, it can be observed that the contribution of the D/G sets overload [H-1-1] sub-hazard to risk is smaller in the system with batteries than in the initial system design. This can be attributed to the fact that batteries act as an additional barrier to the overload sub-hazard. However, despite this, the total fire risk (including H-2 and H-5 hazards) as can be observed is significantly higher in the system with batteries, as the batteries themselves are a new potential source of fire.

Comparing Figure 6 with Figure 7, it can be observed that the relative contribution to the total risk of the UCAs related to [H-1-1] sub-hazard (48%) is double of the relative contribution of the UCAs number associated with [H-1-1] sub-hazard to the total (24% in the initial design). Similarly, the number of the UCAs contributing to H-1-4 sub-hazard is 34% of the total contribution number, yet their risk is only 11% of the total. This is due to the abundancy of barriers tackling the problem of the D/G sets unavailability (sub-hazard [H-1-4]), compared to the other hazards, such as redundancy in available D/G sets, whilst D/G set overload condition (sub-hazard [H-1-1]) can lead to a hazardous condition if few barriers are faulty. Therefore, the scenario number can be considered as inappropriate metric for safety comparison of different systems.

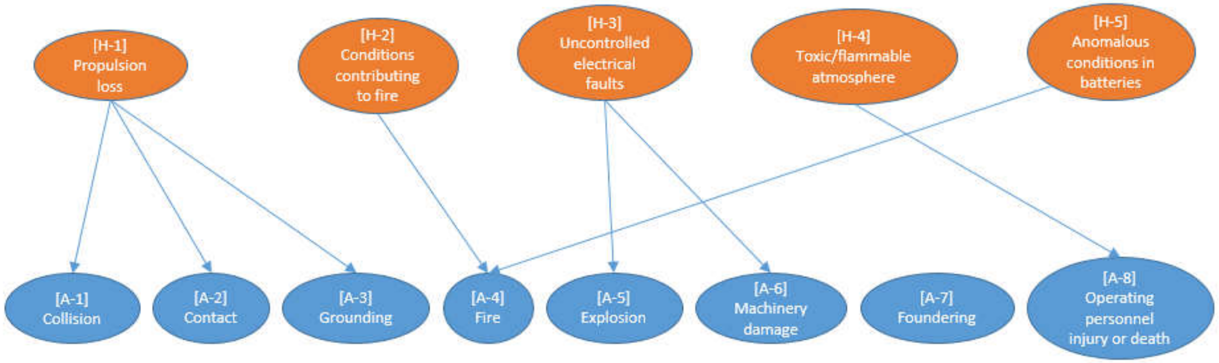

[H-1] Propulsion loss
[H-2] Conditions contributing to fire
[H-3] Uncontrolled electrical faults
[H-4] Toxic/flammable atmosphere
[H-5] Anomalous conditions in batteries
[A-1] Collision
[A-2] Contact
[A-3] Grounding
[A-4] Fire
[A-5] Explosion
[A-6] Machinery damage
[A-7] Foundering
[A-8] Operating personnel injury or death

Figure 5 Interconnection between hazards and accidents.

Table 6 Example of application of the method.

Controller	Controlee	Control action	Failure mode	Context	Assumption	Hazard /Sub hazard	Accident	Causal factor	Mitigating barriers	Environmental factors	m	UC_k	FI_{cf}	M_k	E_k	SI_k	Risk
Fire Fighting control	DG sets	Disconnect energy supply	Providing causes hazards	Power demand status is HIGH and Operating status is ENGAGED	Loss of power generation for several D/G sets simultaneously	H-1-1	Collision/ Contact/ Grounding [A-1],[A-2], [A-3]	Wrong software rules	A) Engine room crew restoring normal provision of fuel to the D/G sets B) Propulsion motors power reduction systems	A) Other vessels in proximity B) Inadequate communication between vessels crew C) Bad weather conditions	4	3	4	3	1	3	4x10^{-7}

Some critical UCAs are provided in Table 7. As it can be observed, failures in the power reduction functions applied during hazardous conditions are considered as the most critical in both systems, as they constitute the last safety barrier before blackout in the systems. Another critical failure is the faulty tripping of the D/G sets by the firefighting system in an engine room, as in this case more than one D/G set can be disconnected from the network, leading to D/G sets overload conditions. In a power system with batteries, the batteries failures management is also considered as critical, as it can lead to fire with a reduced mitigation measures number. Hence, proper design and testing of these functions shall be ensured in the power system.

Table 7 Comparison between initial and system with batteries.

STPA results	Initial design	Batteries included
UCA number	215	300 (+40%)
Causal factors number	2247	3228 (+43%)
Estimated risk PLL [fatalities/year]	6.19 10^{-7} (SI=2) – 6.19 10^{-6} (SI=3)	7.17 10^{-7} (SI=2) – 7.17 10^{-6} (SI=3) (+16%)
Estimated risk IR [fatalities/year]	1.03 10^{-8} (SI=2) – 1.03 10^{-7} (SI=3)	1.20 10^{-8} (SI=2) – 1.20 10^{-7} (SI=3)
Sample of most critical UCAs	- Firefighting system falsely activates quick closing fuel valve - Power Management System (PMS) disconnects consumers necessary for power generation functions, during overload conditions - PMS falsely reduces the propulsion motors and bow thrusters speed (and hence load) - PMS trying to disconnect the already disconnected heavy consumers, hence not allowing the implementation of power reduction function on propulsion motors and thrusters. - PMS failing to reduce thrusters load	- Battery management system not disconnecting the batteries from the network during electrical fault - Battery management system not increasing the cooling during electrical fault conditions. - Firefighting system falsely activates quick closing fuel valve - PMS falsely reduces the propulsion motors and the bow thrusters speed (and hence load) - PMS trying to disconnect the already disconnected heavy consumers, hence not allowing the implementation of power reduction function on propulsion motors and the thrusters.

Table 8 Distribution of risks for initial and system with batteries.

Controller	Initial PLL	Hybrid PLL
AVR	4.80E-07	4.80E-07
BMS	0.00E+00	1.90E-06
Bus-tie controller	1.10E-07	1.10E-07
ECU 7 controller	4.53E-07	3.41E-07
EIM controller	3.57E-07	1.30E-07
Firefighting controller	1.08E-06	1.08E-06
GCU	1.08E-06	9.67E-07
PMS	2.62E-06	2.15E-06
Sea Water Cooling Pump controller	1.60E-08	1.42E-08
Thermostat	1.60E-09	1.42E-09
Total	6.19E-06	7.17E-06

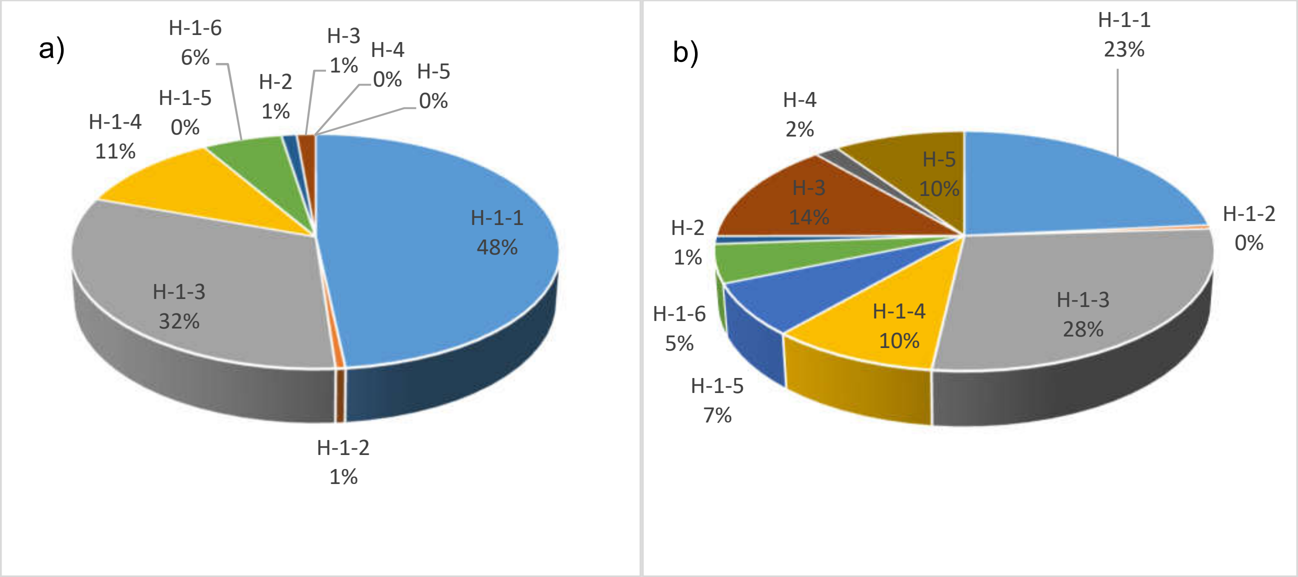


Figure 6 Distribution of estimated risk per hazards a) for initial power system b) for power system with batteries.

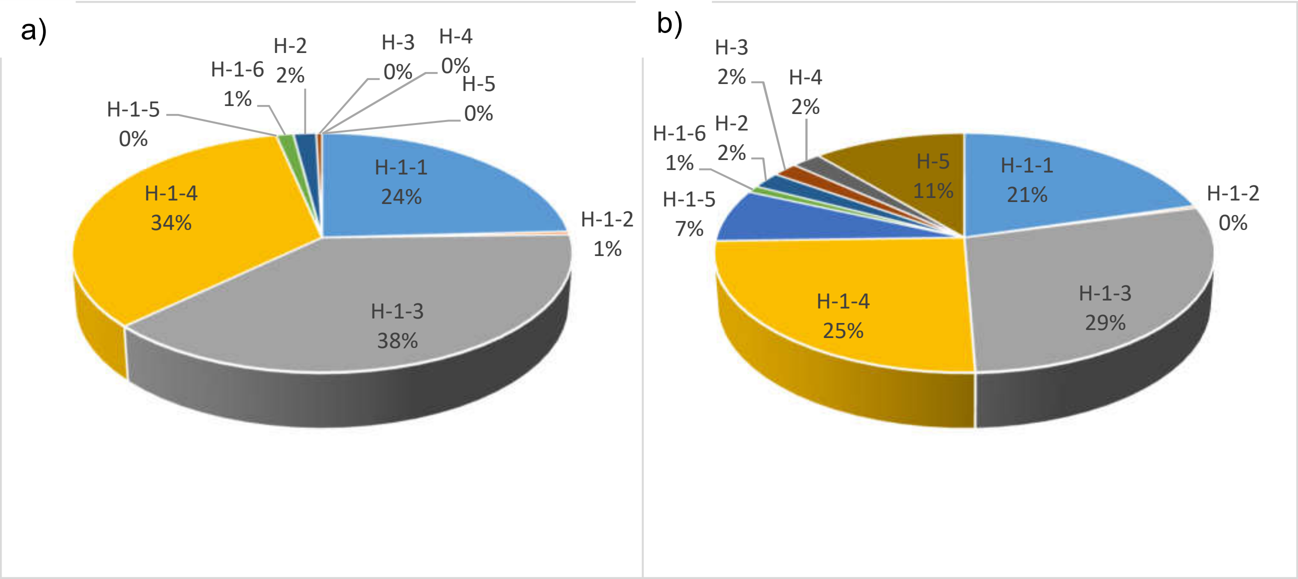


Figure 7 Distribution of identified UCAs per hazard a) for the initial system b) for the system with batteries.

As it can be observed from the results, the method allowed a rough estimation of the risk metrics for different hazardous scenarios, the overall risk for the system and comparison of risk for different systems. It was also possible to estimate the risk for different hazards and controllers. Furthermore, the most critical controllers and scenarios in each system were highlighted. However the estimated risk was not for the whole ship but for a specific system which complicated the comparison with IMO acceptable values. In additions for the system risk estimation, some failure driven scenarios have not been included. Further guidance on how to estimate the UCA consequences and inadvertent environmental factors probability would be also beneficial for this approach. Last, but not least there are several numerical approximations in the methods.

5 CONCLUSIONS

In this study, a new approach for estimating risk metrics in a system based on the STPA has been presented. The proposed approach was applied for comparison of Direct Current power system with Direct Current power system with batteries on an SOV vessel.

The main findings of this study can be summarised as follows:

- The new method allowed risk metrics estimation and comparison for different systems as well as ranking of different scenarios.
- The estimated risk for the failures in control systems, for both systems, is in tolerable regions, according to criteria set by the method.
- The risk, in the power system with batteries may slightly increase due to the increase in the number of scenarios leading to fire
- The risk due to D/G sets overload reduces in system with batteries as batteries act as an additional barrier to the propulsion loss hazard.
- Comparing the number of hazardous scenarios for two systems can lead to wrong conclusions. Still the hazardous scenarios number can be used for comparison of systems complexity.
- The new approach can be used as basis for development of a method for safety comparison between cyber-physical systems.

Whilst the applied methodology was useful for identifying the critical UCAs and comparing risk metrics failures for different systems, still it can be considered as a premature. The methodology could be enhanced by incorporating uncertainty analysis or by integrating it with other methods. The approach could also be enhanced by incorporating multiple experts ranking. However, all these constitute suggestions for future research.

ACKNOWLEDGEMENTS

The work described in this paper was produced in research project NEXUS[†]. The project has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 774519. Kongsberg Maritime CM AS is kindly acknowledged for provision of the relevant system information. The authors greatly acknowledge the funding from DNV GL AS and RCCL for the MSRC establishment and operation. The opinions expressed herein are those of the authors and should not be construed to reflect the views of EU, DNV GL AS and RCCL.

REFERENCES

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research, 253*(1), 1-13. Retrieved from http://www.sciencedirect.com/science/article/pii/S0377221715011479
https://ac.els-cdn.com/S0377221715011479/1-s2.0-S0377221715011479-main.pdf?_tid=2806d80e-082b-11e8-b387-00000aacb35d&acdnat=1517584388_3d54b30e235ab5523e73a989bd786ac4. doi:http://dx.doi.org/10.1016/j.ejor.2015.12.023

Becker, C., & Van Eikema Hommes, Q. (2014). *Transportation systems safety hazard analysis tool (SafetyHAT) user guide (version 1.0)*. Retrieved from

Blandine, A. (2013). *System theoretic hazard analysis applied to the risk review of complex systems: an example from the medical device industry.* (Doctor of Philosophy), Massachusetts Institute of Technology, Cambridge, MA, USA Retrieved from https://dspace.mit.edu/handle/1721.1/79424 (849655099)

Bolbot, V., Theotokatos, G., Boulougouris, E., & Vassalos, D. (2019). *Comparison of diesel-electric with hybrid-electric propulsion system safety using System-Theoretic Process*

[†] www.nexus-project.eu

Analysis. Paper presented at the Propulsion and Power Alternatives, London, United Kingdom.

Bolbot, V., Theotokatos, G., Bujorianu, L. M., Boulougouris, E., & Vassalos, D. (2019). Vulnerabilities and safety assurance methods in Cyber-Physical Systems: A comprehensive review. *Reliability Engineering & System Safety, 182*, 179-193. Retrieved from http://www.sciencedirect.com/science/article/pii/S0951832018302709. doi:https://doi.org/10.1016/j.ress.2018.09.004

Bolbot, V., Theotokatos, G., & Vassalos, D. (2018). *Using system-theoretic process analysis and event tree analysis for creation of a fault tree of blackout in the Diesel-Electric Propulsion system of a cruise ship*. Paper presented at the International Marine Design Conference XIII, Helsinki, Finland.

Brandsaeter, A., Valoen, L. O., Mollestad, E., & Haugom, G. P. (2015). In focus – the future is hybrid. *DNV GL*. Retrieved from www.dnvgl.com/maritime/advisory/battery-hybrid-ship-service.html

BSI. (2004). Functional safety - Safety instrumented systems for the process industry sector. In *Part 3: Guidance for determination of the required safety integrity levels* (Vol. IEC-61511).

Dawson, L. A., Muna, A. B., Wheeler, T. A., Turner, P. L., Wyss, G. D., & Gibson, M. E. (2015). *Assessment of the Utility and Efficacy of Hazard Analysis Methods for the Prioritization of Critical Digital Assets for Nuclear Power Cyber Security*. Retrieved from https://www.osti.gov/servlets/purl/1252915

DNVGL. (2015). Dynamic positioning vessel design philosophy guidelines. Recommended practice (DNVGL-RP-E306). In.

EC. (2019). NEXUS - Towards Game-changer Service Operation Vessels for Offshore Windfarms. Retrieved from https://ec.europa.eu/inea/en/horizon-2020/projects/h2020-transport/blue-growth/nexus

Geertsma, R. D., Negenborn, R. R., Visser, K., & Hopman, J. J. (2017). Design and control of hybrid power and propulsion systems for smart ships: A review of developments. *Applied Energy, 194*, 30-54. Retrieved from http://www.sciencedirect.com/science/article/pii/S0306261917301940
https://ac.els-cdn.com/S0306261917301940/1-s2.0-S0306261917301940-main.pdf?_tid=c25a54a8-082b-11e8-8fb4-00000aab0f26&acdnat=1517584647_0166d0b4d7d583733c6775031f16cdae. doi:http://doi.org/10.1016/j.apenergy.2017.02.060

Goerlandt, F., Khakzad, N., & Reniers, G. (2016). Validity and validation of safety-related quantitative risk analysis: A review. *Safety Science, 99*(November), 127-139. Retrieved from http://www.sciencedirect.com/science/article/pii/S0925753516301795
https://ac.els-cdn.com/S0925753516301795/1-s2.0-S0925753516301795-main.pdf?_tid=c7347efe-082b-11e8-981d-00000aacb35e&acdnat=1517584655_29fa2fde71b875e9cbee7dbeadd193b4. doi:http://dx.doi.org/10.1016/j.ssci.2016.08.023

Hamann, R., Papanikolaou, A., Eliopoulou, E., & Golyshev, P. (2013). Assessment of safety performance of container ships. *Proceedings of the IDFS*, 18-26.

Hill, D. M., Agarwal, A., & Gully, B. (2015). A review of engineering and safety considerations for hybrid power (Lithium-Ion) systems in offshore applications. *Oil and Gas facilities, June 2015*, 68-77.

IMCA. (2015). International Guidelines for The Safe Operation of Dynamically Positioned Offshore Supply Vessels (182 MSF Rev. 2). In.

IMO. (2008). *Formal Safety Assessment - Cruise ships*. Retrieved from

International Maritime Organisation. (2013). *Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process*. London Retrieved from http://research.dnv.com/skj/IMO/MSC-MEPC%202_Circ%2012%20FSA%20Guidelines%20Rev%20III.pdf

Jeong, B., Oguz, E., Wang, H., & Zhou, P. (2018). Multi-criteria decision-making for marine propulsion: Hybrid, diesel electric and diesel mechanical systems from cost-environment-risk perspectives. *Applied Energy, 230*, 1065-1081. Retrieved from http://www.sciencedirect.com/science/article/pii/S0306261918313850. doi:https://doi.org/10.1016/j.apenergy.2018.09.074

Johansen, I., & Rausand, M. (2012). *Risk metrics: Interpretation and choice.* Paper presented at the Industrial Engineering and Engineering Management (IEEM), 2012 IEEE International Conference on.

Leveson, N. (2011). *Engineering a safer world: Systems thinking applied to safety*: MIT press.

Leveson, N., & Thomas, J. (2018). STPA Handbook. In.

Leveson, N. G. (2011). *Engineering a safer world: Systems thinking applied to safety*. London, England: The MIT press.

Organization, I. M. (2014). *SOLAS: consolidated text of the International Convention of Safety of Life at Sea, 1974, as amended* (6th consolidated edition ed.): International Maritime Organization.

Puisa, R., Lin, L., Bolbot, V., & Vassalos, D. (2018). Unravelling causal factors of maritime incidents and accidents. *Safety Science, 110*, 124-141. Retrieved from http://www.sciencedirect.com/science/article/pii/S0925753518304545. doi:https://doi.org/10.1016/j.ssci.2018.08.001

Räsänen, J.-E. (2017). *Current and future scale limitation for alternative marine power and propulsion solutions*. Paper presented at the Power & Propulsion Alternatives for Ships, Rotterdam, Netherlands.

Rokseth, B., Utne, I. B., & Vinnem, J. E. (2017). A systems approach to risk analysis of maritime operations. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 231*(1), 53-68. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1748006X16682606
http://journals.sagepub.com/doi/pdf/10.1177/1748006X16682606. doi:doi:10.1177/1748006X16682606

Sulaman, S. M., Beer, A., Felderer, M., & Höst, M. (2017). Comparison of the FMEA and STPA safety analysis methods–a case study. *Software Quality Journal*, 1-39.

Thomas, J. (2013). *Extending and automating a systems-theoretic hazard analysis for requirements generation and analysis.* Massachusetts Institute of Technology,

United Kingdom Protection & Indemnity Club. (2015). *Risk Focus: Loss of power*. Retrieved from

Vassalos, D., Atzampos, G., Paterson, D., Cichowicz, J., Bertheussen Karolius, K., Boulougouris, E., & Konovessis, D. (2019). Intact stability of passenger ships: safety issue or design concern? Neither!

Vinnem, J.-E. (2014). *Offshore Risk Assessment vol 1*: Springer.

ABRREVIATIONS LIST

AVR	Automatic Voltage Regulator
BMS	Battery Management System
BTU	Bus Tie Unit
D/G	Diesel Generator
DC	Direct Current
FMEA	Failure Modes and Effects Analysis
IMO	International Maritime Organisation
SOV	Service Operation Vessels
STPA	System-Theoretic Process Analysis
PMS	Power Management System
PLL	Potential Loss of Life
UCA	Unsafe Control Actions

Changing risks in existing gas infrastructure in the Netherlands: are traditional hazard analysis methods equipped for an energy transition?

Ben Riemersma[1]

Delft University of Technology (Values of Technology & Innovation, The Netherlands)

ABSTRACT

Countries with extensive gas infrastructures are increasingly turning towards gasses that are produced from renewable energy sources, such as biomass, solar and wind. While these renewable gasses such as biogas/green gas and possibly hydrogen are compatible with existing infrastructure, they exhibit different combustion and explosion behavior. Current safety practices designed for natural gas are not sufficient to ensure a similar level of safety, and must be updated to mitigate changing risks. Additionally, new actors are emerging who are involved with the production and distribution process. The current paper analyzes the extent to which the gas sector in the Netherlands is equipped to deal with a changing risk profile by elaborating on two risk analysis methods. These methods are applied to a segment of the green gas. We find that the Bowtie method that is currently used in the sector provides an understanding of the physical and technical aspects of risks related to green gas provision and is instrumental in communicating them to a general audience. It is also, however, largely static and does nog accommodate changing technical and institutional features of gas provision. The System-Theoretic Accident Model and Processes (STAMP) model, conversely, provides better tools to understand the interaction between incumbent and new actors and technology in the gas sector and provides comprehensive design recommendations for renewable gas systems to a specific audience.

Keywords: STAMP; Bowtie; Renewable Gas; Energy Transition; Safety

1. INTRODUCTION

The transition towards renewable energy sources gives rise to a variety of safety concerns in the Dutch gas infrastructure. Natural gas is increasingly substituted by renewable gasses to curb CO_2 emissions, limit global warming and make up for dwindling domestic natural gas resources. While these renewable gasses can be transported through existing natural gas infrastructure, they exhibit different combustion and explosion behavior compared to natural gas. As is the case for natural gas, the major safety hazard related to renewable gasses is leakage. Across the whole of the gas system, leakage can lead to a wide range of accidents: it can result in poisoning, explosion or fire inside dwellings or outdoors. Yet, the risks posed by renewable gasses—while these include hydrogen and synthetic gasses, we limit these to biogas in the current paper—change and their transport and combustion may also give rise to new hazards. For example, the relative high density of biogas (compared to air, but also compared to natural gas) causes explosions to stay low

---

[1] +31 15 27 89214| b.riemersma@tudelft.nl

above the ground, whereas natural gas explosions shoot up. This may lead to larger damages and requires new or updated mitigation strategies for grid operators and emergency services. However, renewable gasses do not by default pose a higher risk than natural gas. The use of hydrogen, for example, would eliminate the danger of carbon monoxide poisoning because its combustion does not release anything but water, as opposed to the potential carbon monoxide release during natural gas and biogas combustion. In short, actors associated with maintaining safety in and around the gas infrastructure must be prepared for a rapidly changing hazard profile.

Safety in the Dutch gas sector has long been maintained by relatively small group of actors. Traditionally these actors comprised of two large oil and gas companies Shell and Exxon-Mobil to produce and sell natural gas; the Dutch state to orchestrate annual gas production mandates and (high-pressure) infrastructure development; and a variety of municipal and provincial governments that owned local gas (low-pressure) distribution companies. Since the liberalization of the gas sector in the 1990s, gas provision has been strictly separated between commercial and non-commercial activities. Gas production, trade and sales are executed by private parties while the transport of gas is operated by system operators owned by municipalities, provinces and/or the state. Importantly, liberalization facilitated entry to the gas market for new gas producers and granted them access to existing infrastructure. The anticipated growth in renewable gas production is set to increase the number of gas producers and is likely to make gas production more diverse. The production of gas is no longer limited to large oil and gas companies, but increasingly involves actors such as operators of waste treatment plants or farms that produce biogas. As a consequence, existing safety protocols and guidelines must be updated to include changing risks, new hazards *and* more actors involved in (potentially) causing and mitigating these. The current paper investigates if the hazard analyses that are currently in use in the Dutch gas sector can sufficiently address these changing technical and institutional realities.

Hazard analysis in the Dutch gas sector is now primarily done by Bowtie analyses (Coteq Netbeheer 2017; Liander 2015). These analyses are not limited to natural gas, but also actively used to anticipate on hazards associated with biogas and hydrogen (Van Eekelen et al. 2012; KIWA 2018; RVO 2016). Bowtie analyses focus on a central event—the hazard—that may ultimately result in a variety of accidents. Barriers are then identified that can prevent the hazard from happening (the left side of the bowtie), or that can mitigate consequences (the right side) (de Ruijter and Guldenmund 2016). The resulting bowtie gives a visual representation of different hazard scenarios that is easy to understand and popular among academics and practitioners. As industries grow increasingly advanced, however, new hazard analysis methods have been developed that include more (possible) causal factors. Authors propagating these hazard analyses, such as Nancy Leveson and Erik Hollnagel, claim that traditional hazard analyses are ill equipped to analyze safety in industries where increasingly more components interact to a common goal, often assisted by advanced software and cognitively challenging human operator skills (Hollnagel and Goteman 2004; Leveson 2011). Various comparisons between traditional hazard analyses and their newer counterparts have indeed yielded different and complementary results (Chatzimichailidou et al. 2018; Leveson 2016; Merrett et al. 2019). The current article is similar in that it puts forward a comparison between the Bowtie method and Leveson's Systems-Theoretic Accident Model and Processes (STAMP) in order to investigate if the Bowtie method is sufficient in analyzing safety in the rapidly changing Dutch gas system. We test this question by comparing the traditional bowtie analysis as employed in the Dutch gas sector with the new STAMP methodology.

The comparison of both models is focused on the injection of biogas/green gas in the regulated distribution grid. This case reflects the changing reality of gas provision in the Netherlands, as it concerns 1) the production of biogas; 2) the relative small scale of production facilities; and 3) the injection of gasses in the low to medium pressure distribution and not in the centrally controlled high pressure transmission grids. Our findings hold implications for the future development of biogas grids but also for hydrogen and other gasses. The comparison of the two methods yields an extensive list of safety requirements that raise fundamental issues for further research but the theoretical implications reach farther. We link the outcomes of the hazard analyses to specific institutional problems that are left for further research. In the following section, we provide the background of hazard analysis theory and introduce relevant terminology; Section 3 describes the biogas case and explains the use of *green gas*, and analyzes it using both methods; Section 4 compares and analyzes the results, discussing their relevance and possibilities for improvement; Section 5 concludes.

2. HAZARD ANALYSIS THEORY

Several key concepts and words must be clarified before discussing various hazard analyses. The focal point of a hazard analysis, a hazard, is “a set of conditions that may lead to an accident or loss” (Riemersma et al., forthcoming; Leveson, 2013). Hazard analyses aim to identify hazards so as to assist avoiding accidents and losses. The likelihood of accidents and losses happening, combined with their severity—or, in other words, the risk—can be estimated in a consequent risk analysis (Aven 2011; Christensen et al. 2003). Risk is expressed as a number, whereas hazard analysis can also be qualitative. Hazard and risk analyses have at least a century old history, and are developing still[2]. This article discusses two hazard analyses: one that is currently in use in the Dutch gas sector (Bowtie), as well as an alternative (STAMP). We first introduce both methods before applying them in Section 3.

2.1. Bowtie Method

The Bowtie method is a consolidation of several models that were developed over the course of the 1960s and 1970s. It is centered around a critical event and resembles a bowtie as shown in Figure 1. This figure shows how different causes on the left side can initiate the event, after which it can lead to accidents with various consequences; the safety barriers are installed to prevent (left side of the bowtie) the hazard or mitigate (right side) accidents (de Ruijter and Guldenmund 2016; Swuste et al. 2016). Barriers are not limited to the bowtie methodology, and have been used to visualize and understand safety protocols by many

---

[2] Paul Swüste and colleagues provided an excellent overview in various installments (Swuste et al. 2016, 2018; Swuste, Van Gulijk, and Zwaard 2010)(Swuste et al. 2016, 2018; Swuste, Van Gulijk, and Zwaard 2010)(Swuste et al. 2016, 2018; Swuste, Van Gulijk, and Zwaard 2010)(Swuste et al. 2016, 2018; Swuste, Van Gulijk, and Zwaard 2010)

more scholars (Reason 1990). They help to make safety measures more visible and identify areas that lack sufficient protection (De Dianous and Fievez 2006).

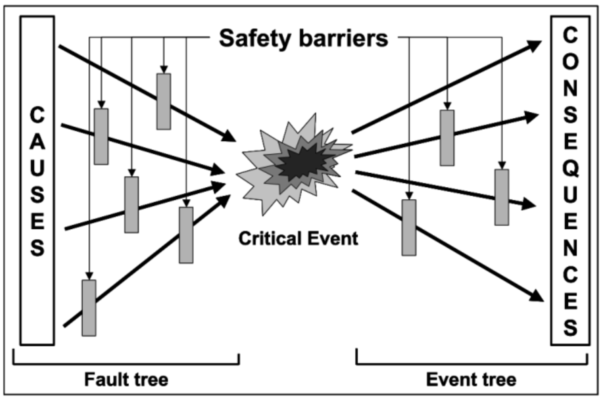


Figure 1. Typical Bowtie analysis (De Dianous and Fievez 2006)

There are a number of varieties of the Bowtie method; the version most common in the gas industry in the Netherlands and also used in this paper was developed by Shell (de Ruijter and Guldenmund 2016). Compared to other variations of the same method, the Shell version is more straightforward and serves to visualize simple cause-effect relationships. Single root causes have a direct line to the central event, and from there can cause any accident outlined on the right hand side (cf. Figure 1). This makes it difficult to analyze how multiple factors can all contribute to an accident. This Shell Bowtie explicitly calls for the level of detail to remain “within reasonable boundaries” so that it remains useful and relevant for its target audience (Visser 1998, 52).

2.2.STAMP Method

There are many methods that provide alternatives to the Bowtie. However, the Bowtie method, as well as other popular hazard analysis methods such as hazard and operability studies (HAZOP) and failure mode and effect analysis (FMEA)[3], have been criticized for not being able to incorporate larger dependencies and more advanced technologies of current systems (Cameron et al. 2017; Dunjó et al. 2010). This section focuses on a hazard analysis tool derived from Nancy Leveson’s System-Theoretic Accident Model and Processes (STAMP) model. This tool, called System-Theoretic Process Analysis (STPA), shifts the focus away from accident or operator failure to including the wider environment in which hazards emerge. The analysis identifies the conditions under which a system is in control, and specifies the requirements that must be met to preserve the safe state. A loss of control—the release of a hazard—results from failure to enforce safety requirements. The STPA approach allows for identification of design errors (either hardware or software) and organizational failures: even if all individual elements of a system work perfectly, STPA can identify hazards that occur from their interaction through poor design.

STAMP is based on systems theory and views a system as interacting *control loops*. These control loops represent the behavior of different system components (human, physical

---

[3] Also used in the Dutch gas sector, but omitted from this paper.

or social) and function by means of *control actions* executed by *controllers*. The system is interpreted as a hierarchical structure, where each level imposes constraints on the level below it. In other words, control actions enforce safety constraints on lower-level components. The analysis aims to identify the conditions under which control actions become unsafe and, by consequence, can result in a hazard. The control actions are derived from a visual representation of the *control structure*, which represents the functional working of the relevant system. The control structure is accompanied by a *hierarchical structure* that illustrates the elements in the system that exercise control over it. The process will be illustrated by means of a green gas feed-in structure in the next section.

3. Case analysis

Our analysis focuses on hazards related to the provision of biogas in the Netherlands. More specifically, we limit our analysis to the injection of *green gas* into the distribution grid. Biogas refers to all gas produced from biomass but can vary substantially in quality. When biogas is upgraded to match the quality standards of natural gas it is referred to as green gas, and virtually similar to natural gas. Green gas is permitted in existing infrastructure and exhibits the same combustion and explosion behavior as natural gas. It is checked for quality before it enters the grid in order to guarantee compatibility with existing infrastructure and appliances. The inadvertent entry of biogas (i.e. not matching natural gas properties) could jeopardize the integrity of the gas infrastructure in a number of ways. For example, it increases the risk of material deterioration due to aggressive physical properties (i.e. highly corrosive and poisonous hydrogen sulfide); changes explosion behavior so that emergency services are unaware of mitigation strategies; and it emits more toxic gasses upon inadvertent release. The combustion of biogas is also potentially hazardous. Appliances in the Netherlands are attuned to a specific gas quality; departing from these specifications may result in incomplete or faulty combustion, resulting in leakage and risk of poisoning (KIWA 2018; RVO 2016). It is therefore essential to supervise green gas production in order to limit grid entry only to gas that meets the requirements set out in national legislation.

While natural gas and green gas are alike in physical properties, their production methods differ significantly. Natural gas is extracted from domestic resources or imported from abroad on a large scale after which is it is transported at high-pressure (67-80 bars) through a *transmission* system connected to large industrial users and points that connect to lower-pressure (4-8 bars) *distribution* systems. At these points, responsibility for safe transport shifts accordingly from a single national *transmission* system operator (TSO) to any of the seven *distribution* system operators (DSO) located throughout the country. DSOs function as a regional monopoly and have a dedicated service area; they transport the gas onwards to industrial customers attached to the 4-8 bar distribution network until gas eventually reaches small industries and houses through a 100-300 millibar network. Quality control for natural gas is executed by the TSO and takes place at the transfer point between transmission and distribution grid. Hence, DSOs have always been ensured of uniform gas quality in their networks. Unlike natural gas, however, green gas is injected directly into the distribution grid.

The emergence of green gas production created new roles for DSOs. They are responsible for the quality of gas injected into their grid, so they perform checks on the gas that is delivered by green gas producers. While the share of green gas production is still small at 0.3% of the total gas consumption, it is growing rapidly. 2018 saw the total production rise by 11% to 109m^3 and the total amount of green gas producers increase from 36 to 43 (Netbeheer Nederland 2019b). In short, the provision of gas—both natural and green—is set to change rapidly. In the next two sections we will analyze hazards associated with green gas feed-in using two methods. The Bowtie analysis (§3.1) is adapted from a variety of Dutch research reports that investigate hazards associated with biogas, green gas and hydrogen. It focuses mostly on the physical properties of biogas, as these are relevant during the process of upgrading to green gas as well as in the case of inadvertent release of biogas into the distribution grid. The STAMP analysis (§3.2) will build on and elaborate the findings of the Bowtie analysis, and add a stronger organizational focus.

3.1. Bowtie Analysis

The Bowtie is currently the most popular hazard analysis method in the Dutch gas sector. Coteq, a DSO that operates in the eastern part of the Netherlands, details their hazard analysis strategy in their yearly report (Coteq Netbeheer 2017). They develop several Bowtie methods in order to visualize risks and safety concerns. The Bowties are periodically updated by assessing new safety concerns, such as those caused by renewable gasses, that are collected through a variety of sources, both internal and external. DSOs collaborate to develop Bowties for gas provision and disseminate relevant information between them.[4] However, DSOs are not the only institutions that make use of this hazard analysis: a report commissioned by the ministry of Economic Affairs also relies on Bowtie methodology to propose guidelines for biogas transport (RVO 2016) and the same holds for an inquiry into the preparedness of Dutch gas networks to renewable gasses (Netbeheer Nederland 2019a).

The analyses quoted above show strong similarities in analyzing hazards associated with renewable gasses. All use established natural gas hazards as a reference point after which hazards associated with biogas or hydrogen are added or modified. The most detailed analysis, concerning biogas hazards, is shown in the appendix and results in a list of barriers summarized in Table 1. Technical consultancy and certification institute KIWA analyzed the hazards associated with hydrogen provision in a similar way (KIWA 2018). They tested the efficiency of barriers in the natural gas bowtie for hydrogen provision, specifying new and changing concerns for safety.

Table 1: Relevant information from Bowtie analysis (summarized from Appendix 1)

Notable barriers (aimed at threat)	Notable barriers (mitigating accident)
Corrosion/deterioration of pipeline due to biogas properties: • Adapt pipeline quality to withstand hazardous biogas properties • Regulate biogas quality so as to limit the amount of hazardous properties • Stop feed-in process when gas leak has occurred	**Fire, explosion, suffocation, intoxication in rural areas** • Stop gas flow by facilitating leak recognition (i.e. detectors, gas smell, awareness of emergency contact info) • Incorporate distance of 3.5 meters between biogas pipeline and buildings and check trajectory yearly for changes
Excavation damage	• Create awareness among emergency

---

[4] Interview with DSO, may 15th 2019

• Organize required permission for excavation work • Additional preventive measures (i.e. guide excavation activities) **Failing grid connections** • Design leak-proof grid connections **Maintenance works on gas network in operation** • Additional preventive measures (i.e. mark biogas pipelines) • Create more awareness concerning hazards related to biogas	services (i.e. fire fighters, police, ambulance) and update contingency plans to include hazards specific to biogas **Gas Leakage in construction site (leading to intoxication)** • Mitigate damage by removing people from danger zone—wear H2S detecting masks • Incorporate personal protection measures against H2S and moist gas

3.2. STPA Analysis

We limit the STPA analysis to the feeding in of green gas to the grid. This part of the renewable gas infrastructure is especially relevant because it reflects the larger diversity of actors associated with future gas provision.

3.2.1. Accidents to consider, high-level hazards and high-level safety constraints

The STAMP analysis starts by listing accidents that must be avoided. Following STAMP terminology we define an accident as “an undesired or unplanned event that results in a loss, including loss of human life or human injury, property damage, environmental pollution, mission loss etc.” (Leveson 2011, 181). A list of accidents to be considered for the transport of biogas was adapted from the Bowties analyzed in the previous section and include:

- A1: Fire/Explosion/Suffocation/Intoxication in rural areas
- A2: Fire/Explosion/Suffocation/Intoxication in urban areas
- A3: Fire/Explosion/Suffocation/Intoxication inside dwellings
- A4: Loss of revenue

As is the case in §3.1, the accidents are all potential results of gas leakage. We specify a number of *high-level safety hazards* at the outset of our analysis with the aim of further specifying them. Similarly, we identify *high-level safety constraints* that are generic safety practices that must be put in place to prevent safety hazards from happening. These high-level safety constraints, too, will be further refined during the analysis. Both are summarized in Table 2.

Table 2: High-level safety hazards and High-level safety constraints

High-level Safety Hazards (HLSH)	High-level Safety Constraints (HLSC)
[HLSH-1] Uncontrolled release on-spec gas from the distribution grid [A-1, A-2]	[HLSC-1] The hierarchical control structure (HCS) must prevent uncontrolled gas release [HLSH-1, HLSH-2]
[HLSH-2] Uncontrolled release out-of-spec gas from the distribution grid [A-1, A-2]	[HLSC-2] The HCS must respond to an uncontrolled gas release so as to minimize its consequences [HLSH-1, HLSH-2]
[HLSH-3] Feeding in of out-of-spec gas into the distribution grid [A-1, A-2, A-3]	[HLSC-3] The HCS prevent out-of-spec gas to be fed

[HLSH-4] Interruption of gas supply [A3]	into the distribution grid [HLSH-3] [HLSC-4] The HCS must attend to safely continue gas supply after any possible interruption [HLSH-4]

3.2.2. Control structure, high-level hierarchical structure and safety control actions

Figure 2 (next page) shows the control structure for a feed-in installation of green gas, as well as the authorities surrounding it in the high-level hierarchical structure. All elements of the hierarchical structure exert influence on the control structure by imposing constraints by means of regulation, instructions or certification for example. These constraints can be further specified according to the information derived from the current analysis.

The control structure—within the dotted line—is our focal part of analysis. It illustrates the process where the gas is transported by the *gas producer* from the production plant to the regulated gas grid. The gas producer has equipment to verify if the produced gas meets the quality as specified by the Gas Act; values of Methane (CH4), Hydrogen Sulfide (H2S), Carbon Dioxide (CO2), Oxygen (O) and Nitrogen (N) are checked every 5 minutes. Based on the feedback received from these sensors, the producer can execute two *control actions:* 1) continue gas production or 2) stop gas production. Before the gas is injected into the gas grid, then, the *distribution system operator* provides another quality check in order to independently verify whether or not the gas meets requirements. The information regarding gas quality is received through sensors and sent to a DSO operating center. From there, the DSO can decide remotely whether to 3) feed-in the gas or 4) eject the gas into the air (i.e. flare it).

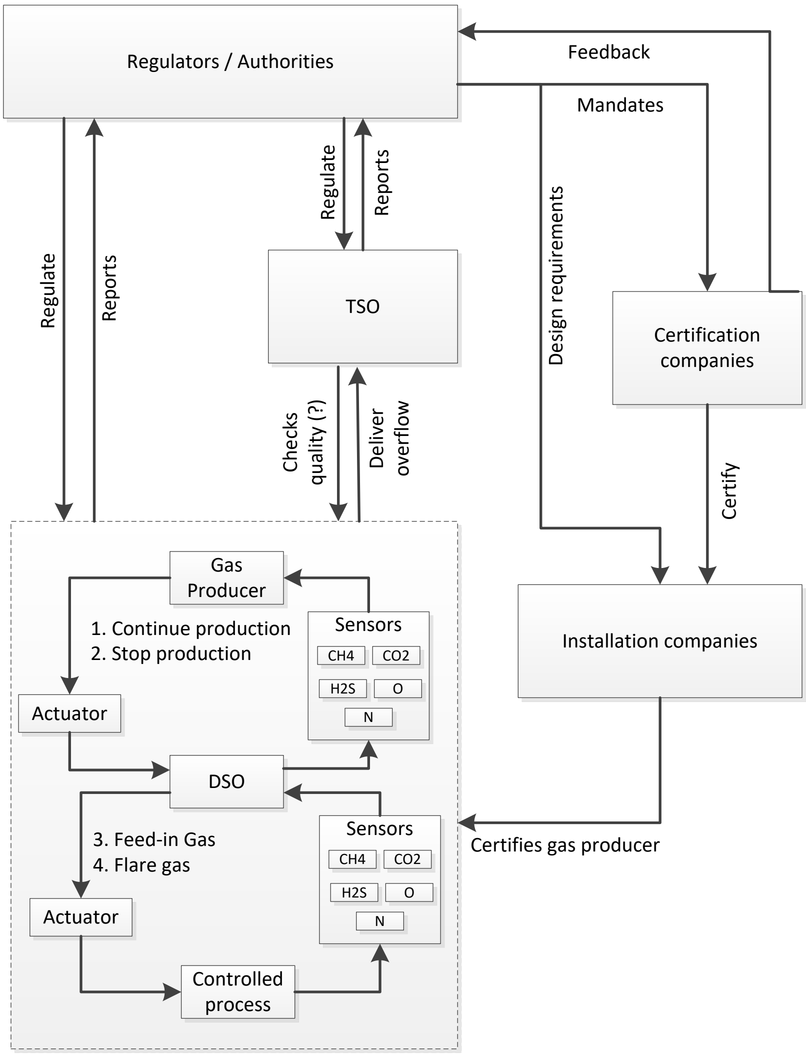


Figure 2: Control structure diagram and high-level hierarchical structure

3.2.3. Potentially unsafe control actions and safety requirements

The previous section identified four control actions. We will now analyze the conditions under which these can become unsafe. Control actions can generally be hazardous in four ways (Leveson 2011):

1. Control action required for safety is not provided or not followed
2. An unsafe control action is provided that leads to a hazard
3. A potentially safe control action is provided too late, too early, or out of sequence
4. A safe control action is stopped too soon or applied too long (for a continuous or non-discrete control action)

For all four identified control actions, we analyze whether any of these four conditions can lead to unsafe behavior. This is summarized in Table 3 for two of the four possible control actions. For our current purposes, we have limited the analyses to control actions 3 and 4. 1 and 2 may follow at a later stage. Control Actions 3 and 4 can yield hazardous situations under 4 specific circumstances—outlined in Table 4 as Unsafe Control Actions (UCAs) 1 through 4. Safety constraints must be designed in order to mitigate these four UCAs; these are also shown in Table 4.

Table 3: identifying unsafe control actions

Control action	Not providing causes hazard	Providing causes hazard	Wrong timing or order causes hazard	Stopped too soon or applied too long
1. Continue production	To Be Determined	TBD	TBD	TBD
2. Stop production	TBD	TBD	TBD	TBD
3. Feed-in gas	Controller does not provide gas feed-in when the gas is on-spec (1)	Controller provides gas feed-in when it is out-of-spec (2)	Controller provides feed-in gas too early when gas is out-of-spec (3)	Controller provides feed-in gas too long when gas is out-of-spec (4) Controller provides feed-in gas too short when gas is on-spec (5)
4. Flare gas	Controller does not provide flaring of gas when gas is out-of-spec (6)	Controller provides flaring of gas when gas is on-spec (7)	Not hazardous	Controller provides flaring of gas too late when gas is out-of-spec (8)

Table 4: Relating unsafe control actions to safety constraints

Unsafe Control Actions (UCA)		Safety constraints (SC)	
UCA-1	Controller does not provide gas feed-in when the gas is on-spec (A4)	**SC-1**	Controller must provide gas feed-in when the gas is on-spec [UCA-1]
UCA-2	Controller provides gas feed-in when it is out-of-spec (A1, A2, A3)	**SC-2**	Controller must not provide gas feed-in when it is out-of-spec [UCA-2]
UCA-3	Controller provides feed-in gas too early when gas is out-of-spec (A1, A2, A3)	**SC-3**	Controller must feed-in gas only if it is on-spec [UCA-3, UCA-4, UCA-5]
UCA-4	Controller provides feed-in gas too long when gas is out-of-spec (A1, A2, A3)	**SC-4**	Controller must provide flaring of gas when gas is out-of-spec [UCA-6]
UCA-5	Controller provides feed-in gas too short when gas is on-spec (A4)	**SC-5**	Controller must not provide flaring when gas is on-spec [UCA-7]
UCA-6	Controller does not provide flaring of gas when gas is out-of-spec (A1, A2, A3)	**SC-6**	Controller must provide flaring only when gas is out-of-spec [UCA-8]
UCA-7	Controller provides flaring of gas when gas is on-spec (A4)		
UCA-8	Controller provides flaring of gas too late when gas is out-of-spec (A1, A2, A3)		

3.2.4. Identifying Loss Scenarios

The identified Unsafe Control Actions and Safety Constraints can be further understood by creating loss scenarios. Creating scenarios possibly enables the identification of more detailed safety constraints and it is also important in identifying how several factors may interact to lead to a hazard. This step, also referred to as STPA Step 2, will be executed for a subsequent version of the current article and is deemed more effective when there is consensus on the appropriate execution of earlier steps (i.e. §3.2.1. – §3.2.3.).

4. ANALYSIS AND DISCUSSION

The two hazard analyses yield significantly different results. The results—summarized in table 5—indicate that the Bowtie yields generic recommendations whereas STPA provides more detailed and in-depth results (i.e. safety constraints*)*. The Bowtie analysis' visual representation allows for a good communication of risk and barriers to a general audience, whereas the STAMP outcomes are better suited for insiders with a high level of specific knowledge. These findings are in line with the purposes of both methods as indicated in Section 2 and also resemble earlier comparisons (Chatzimichailidou et al. 2018; Merrett et al. 2019).

Table 5: Comparison Bowtie and STPA analysis

Bowtie method results	STPA method results
• Regulate biogas quality so as to limit the amount of hazardous properties • Stop feed-in process when gas leak has occurred	• Controller must provide gas feed-in when the gas is on-spec • Controller must not provide gas feed-in when it is out-of-spec • Controller must feed-in gas only if it is on-spec • Controller must provide flaring of gas when gas is out-of-spec • Controller must not provide flaring when gas is on-spec • Controller must provide flaring only when gas is out-of-spec

The wider scope of the STPA analysis allows for a more detailed analysis of complex systems. The results of a more detailed analysis are already visible in Table 5. Unlike the results of the Bowtie methodology, the six Safety Constraints generated from the STPA method pay attention to specific conditions that may contribute to accidents. Not only does STPA generate safety constraints that target physical damage, it is worth noticing that the Safety Constraint [SC-5] targets loss of revenue [A4]—a kind of accident unlikely to be covered in more conventional safety analyses. The STPA method captures more links in the chain leading to possible accidents: where the Bowtie method links action recommendations to a gas leak (i.e. a *hazard*), the STPA method links them to the measurement of gas quality (i.e. a *condition* that may result in a hazardous state when out-of-spec).

The list of Safety Constraints that results from the STPA analysis raises many questions with regards to their preferred implementation. Most immediately, we have not distinguished among the two controllers currently involved with green gas provision: distribution system operators and green gas producers. Both are faced with new roles that must be embedded in existing institutional structures. The governance of these existing structures must be modified to allocate more responsibilities to said actors and facilitate communication and feedback among them, as well as incumbents such as lawmakers, regulators and the transmission system operator. New coordination problems emerge with regards to essential responsibilities for safety governance such as quality control for green gas injected into the transmission grid or the increased involvement of publicly owned DSOs in commercial tasks. These issues must be addressed in further research, but are taken into consideration for future versions of the current paper.

The data for both hazard analyses can be extended by further feedback from relevant actors to arrive at stronger outcomes. The current data has been gathered from interviews with relevant stakeholders over the past two years as well as extensive literature review (Riemersma, Correljé, and Künneke 2019). Further interviews and site visits should be conducted focusing on the interaction between the DSO and green gas producers. These should result in more clearly specified control actions, as well as the conditions under which they are rendered unsafe. Potentially, the analysis could be broadened to include the transmission system operator. It is likely that green gas will eventually be transported through high-pressure transmission grids to balance supply and demand. This would introduce the TSO as a *third* relevant controller, further complicating the distribution of safe control actions. Additionally, safety hazards related to the provisioning of green gas may also resurface for hydrogen and other gasses that will increasingly be transported in the Dutch gas system.

5. CONCLUSION

This paper analyzes the safety of renewable gas systems by applying Bowtie and STPA hazard analysis methods. By focusing on the feeding in of green gas into the distribution grid this paper highlights the way in which both hazard analyses emphasize different aspects of safety. The Bowtie method provides a comprehensive but non-exhaustive overview of changing risks compared to the provision of natural gas. The main outcomes relevant to our case include the regulation of biogas/green gas quality to limit the amount of hazardous properties in the gas and stopping the feed-in process once a leak has occurred. It is striking that the Bowtie analysis includes those risks that were present for natural gas (i.e. regulate gas quality), but fails to mention new risks associated with the provision of green gas (i.e. checking for quality of green gas). While these changing risks can be included into an updated Bowtie rather simply, it does underscore the importance of extending the scope of hazard analysis beyond risks traditionally assumed with natural gas distribution.

The STPA focuses on the detection of gas quality before grid entry and develops strategies to prevent the feed in of gas that does not meet quality requirements. A detailed list of Safety Constraints can assist in shaping institutions that effectively govern safety in future gas systems. We illustrate how the STPA is able to capture hazards that did not surface in existing Bowtie analyses, and argue that the method is superior in identifying and illustrating hazards in an increasingly complex gas system. The visualization of the gas system using control loops situated in a hierarchical system provides a clear overview of the different actors at play, and how they relate to the provision of renewable gas. Especially as the Dutch gas system grows increasingly heterogeneous with more and different controllers, the detailed STPA yields better results than the static Bowtie analysis.

The comparison between the two analyses highlights their different focus. The Bowtie analysis is appropriate for educative purposes among academics and particularly practitioners. Its visualization facilitates an easy understanding of causal relationships between accident causes and accidents as well as offers a good template for identifying and strengthening safety barriers. The STPA analysis is more thorough than the Bowtie and yields more detailed and even new results. It is therefore useful for identifying root causes not (currently) present in the Bowtie analysis and valuable for shaping safety policy for renewable gas provision. These results hold implications not only for green gas or biogas. Keeping within gas provision, many of these findings might hold for the future distribution of hydrogen through existing pipelines or even in independently operated natural gas grids. Beyond the gas sector, these results are relevant for water and electricity provision—traditionally centralized service infrastructures which are increasingly disrupted by new actors operating on a decentralized level.

6. ACKNOWLEDGEMENTS

This research is funded by Netbeheer Nederland.

References

Aven, Terje. 2011. “On the New ISO Guide on Risk Management Terminology.” *Reliability Engineering & System Safety* 96(7): 719–26. https://linkinghub.elsevier.com/retrieve/pii/S0951832010002760 (June 4, 2019).

Cameron, Ian et al. 2017. “Process Hazard Analysis, Hazard Identification and Scenario Definition: Are the Conventional Tools Sufficient, or Should and Can We Do Much Better?” *Process Safety and Environmental Protection* 110: 53–70. http://dx.doi.org/10.1016/j.psep.2017.01.025.

Chatzimichailidou, Maria Mikela, James Ward, Tim Horberry, and P John Clarkson. 2018. “A Comparison of the Bow-Tie and STAMP Approaches to Reduce the Risk of Surgical Instrument Retention.” *Risk Analysis* 38(5). https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.12897 (June 27, 2019).

Christensen, Frans Møller, Ole Andersen, Nijs Jan Duijm, and Poul Harremoës. 2003. “Risk Terminology—a Platform for Common Understanding and Better Communication.” *Journal of Hazardous Materials* 103(3): 181–203. https://linkinghub.elsevier.com/retrieve/pii/S0304389403000396 (June 4, 2019).

Coteq Netbeheer. 2017. *Kwaliteits- En Capaciteitsdocument 2017*. Almelo. https://www.coteqnetbeheer.nl/Portals/692/KCD-2017-Coteq-Netbeheer-20171201.pdf (July 15, 2019).

De Dianous, Valérie;, and Cécile Fievez. 2006. “ARAMIS Project : A More Explicit Demonstration of Risk Control through the Use of Bow-Tie Diagrams and the Evaluation of Safety Barrier Performance.” *Journal of Hazardous Materials* 130(3): 220–33. https://hal-ineris.archives-ouvertes.fr/ineris-00961900 (July 17, 2019).

Dunjó, Jordi, Vasilis Fthenakis, Juan A. Vílchez, and Josep Arnaldos. 2010. “Hazard and Operability (HAZOP) Analysis. A Literature Review.” *Journal of Hazardous Materials* 173(1–3): 19–32. https://linkinghub.elsevier.com/retrieve/pii/S0304389409013727 (July 16, 2019).

Van Eekelen, R N, E A Polman, H A Ophoff, and M Van Der Laan. 2012. *New Networks for Biogas - Current Practices, Risks, Cost Aspects*. Apeldoorn. www.kiwagastechnology.nl (March 16, 2018).

Hollnagel, Erik, and Örjan Goteman. 2004. “The Functional Resonance Accident Model.” In *Proceedings of Cognitive System Engineering in Process Plant*, , 155–61. https://s3.amazonaws.com/academia.edu.documents/39790175/403.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1507132658&Signature=7vkuZAbOBO3Gx3qK5fXIvYcZn7s%3D&response-content-disposition=inline%3B filename%3DThe_functional_resonance_accident_model.pd (October 4, 2017).

KIWA. 2018. *Toekomstbestendige Gasdistributienetten*. Apeldoorn.

Leveson, Nancy G. 2011. *Engineering A Safer World: Systems Thinking Applied To Safety*. Cambridge: The MIT Press.

———. 2013. “Engineering a Safer World.” In *Engeneering A Safer World*,.

———. 2016. *CAST Analysis of the Shell Moerdijk Accident*. http://sunnyday.mit.edu/shell-moerdijk-cast.pdf (May 13, 2019).

Liander. 2015. “Kwaliteits- En Capaciteitsdocument Gas 2015.” : 69. https://www.liander.nl/sites/default/files/Kwaliteits-_en_capaciteitsdocument_2015_Gas.pdf.

Merrett, Hew Cameron et al. 2019. “Comparison of STPA and Bow-Tie Method Outcomes in the Development and Testing of an Automated Water Quality Management System.” *MATEC Web of Conferences2* 273. https://doi.org/10.1051/matecconf/201927302008 (June 26, 2019).

Netbeheer Nederland. 2019a. *Betrouwbaarheid van Gasdistributienetten in Nederland 2018*. Den Haag. https://www.netbeheernederland.nl/_upload/Files/Betrouwbaarheid_gasdistributienetten_in_Nederland_2018_151.pdf (July 15, 2019).

———. 2019b. “Steeds Meer Groen Gas in Het Net.” *Online publication*. https://www.netbeheernederland.nl/nieuws/steeds-meer-groen-gas-in-het-net-1271 (July 22, 2019).

Reason, J. T. 1990. *Human Error*. Cambridge: Cambridge University Press.

Riemersma, Ben, Aad Ferdinand Correljé, and Rolf Künneke. 2019. “Historical Developments in the Dutch Gas Sector: Unravelling Safety Concerns in Gas Provision.” *Safety Science* In press.

de Ruijter, A., and F. Guldenmund. 2016. “The Bowtie Method: A Review.” *Safety Science* 88: 211–18. https://linkinghub.elsevier.com/retrieve/pii/S0925753516300078 (July 15, 2019).

RVO. 2016. *Voorstel Voor Richtlijn Voor Het Transport van Ruw Biogas*. Apeldoorn. https://www.rvo.nl/sites/default/files/2016/03/Voorstel voor richtlijn voorhet transport van ruw biogas.pdf (March 29, 2018).

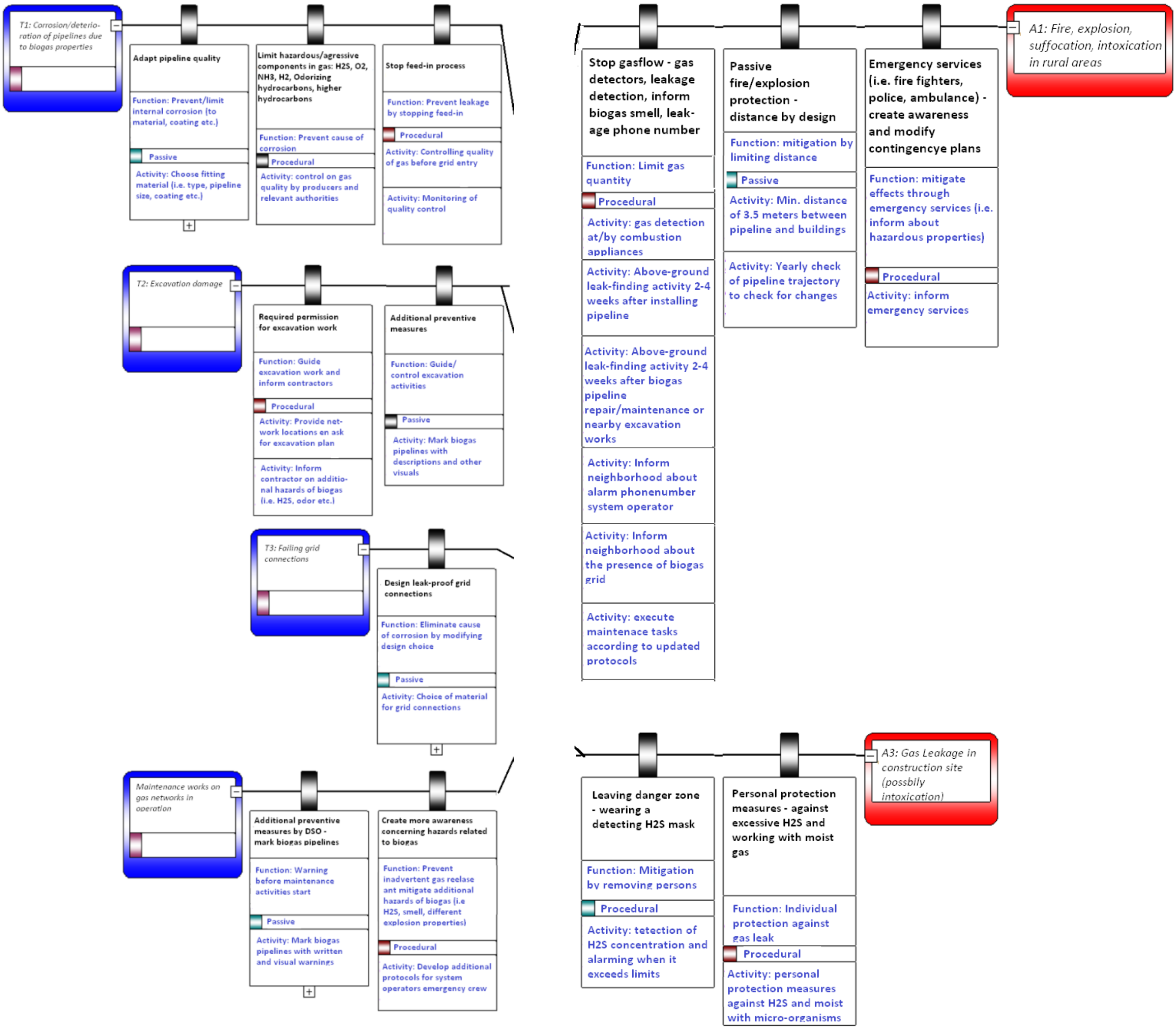
Swuste, Paul et al. 2016. “Developments in the Safety Science Domain, in the Fields of General and Safety Management between 1970 and 1979, the Year of the near Disaster on Three Mile Island, a Literature Review.” *Safety Science* 86: 10–26. http://ac.els-cdn.com/S0925753516000400/1-s2.0-S0925753516000400-main.pdf?_tid=d6194bfe-54e7-11e7-a0e1-00000aab0f01&acdnat=1497874216_ff0ba401775df994a62292464011d712 (June 19, 2017).

———. 2018. “Safety Management Systems from Three Mile Island to Piper Alpha, a Review in English and Dutch Literature for the Period 1979 to 1988.” *Safety Science* 107: 224–44. https://linkinghub.elsevier.com/retrieve/pii/S0925753516304544 (July 16, 2019).

Swuste, Paul, Coen Van Gulijk, and Walter Zwaard. 2010. “Safety Metaphors and Theories, a Review of the Occupational Safety Literature of the US, UK and The Netherlands, till the First Part of the 20th Century.” *Safety Science* 48: 1000–1018. https://tudelft.openresearch.net/image/2016/11/24/swuste_vangulijk_zwaard_2010_safety_science_48.pdf (May 27, 2019).

Visser, J. P. 1998. “Developments in HSE Management In Oil And Gas Exploration And Production.” In *Safety Management: The Challenge of Change*, eds. Andrew Hale and Michael Baram. Oxford: Pergamom, 43–67.

Appendix 1: Edited version of bowtie


T1: Corrosion/deterioration of pipelines due to biogas properties
Adapt pipeline quality
Function: Prevent/limit internal corrosion (to material, coating etc.)
Passive
Activity: Choose fitting material (i.e. type, pipeline size, coating etc.)
Limit hazardous/agressive components in gas: H2S, O2, NH3, H2, Odorizing hydrocarbons, higher hydrocarbons
Function: Prevent cause of corrosion
Procedural
Activity: control on gas quality by producers and relevant authorities
Stop feed-in process
Function: Prevent leakage by stopping feed-in
Procedural
Activity: Controlling quality of gas before grid entry
Activity: Monitoring of quality control
T2: Excavation damage
Required permission for excavation work
Function: Guide excavation work and inform contractors
Procedural
Activity: Provide network locations en ask for excavation plan
Activity: Inform contractor on additional hazards of biogas (i.e. H2S, odor etc.)
Additional preventive measures
Function: Guide/ control excavation activities
Passive
Activity: Mark biogas pipelines with descriptions and other visuals
T3: Failing grid connections
Design leak-proof grid connections
Function: Eliminate cause of corrosion by modifying design choice
Passive
Activity: Choice of material for grid connections
Maintenance works on gas networks in operation
Additional preventive measures by DSO - mark biogas pipelines
Function: Warning before maintenance activities start
Passive
Activity: Mark biogas pipelines with written and visual warnings
Create more awareness concerning hazards related to biogas
Function: Prevent inadvertent gas reelase ant mitigate additional hazards of biogas (i.e H2S, smell, different explosion properties)
Procedural
Activity: Develop additional protocols for system operators emergency crew
Stop gasflow - gas detectors, leakage detection, inform biogas smell, leakage phone number
Function: Limit gas quantity
Procedural
Activity: gas detection at/by combustion appliances
Activity: Above-ground leak-finding activity 2-4 weeks after installing pipeline
Activity: Above-ground leak-finding activity 2-4 weeks after biogas pipeline repair/maintenance or nearby excavation works
Activity: Inform neighborhood about alarm phonenumber system operator
Activity: Inform neighborhood about the presence of biogas grid
Activity: execute maintenance tasks according to updated protocols
Passive fire/explosion protection - distance by design
Function: mitigation by limiting distance
Passive
Activity: Min. distance of 3.5 meters between pipeline and buildings
Activity: Yearly check of pipeline trajectory to check for changes
Emergency services (i.e. fire fighters, police, ambulance) - create awareness and modify contingencye plans
Function: mitigate effects through emergency services (i.e. inform about hazardous properties)
Procedural
Activity: inform emergency services
A1: Fire, explosion, suffocation, intoxication in rural areas
Leaving danger zone - wearing a detecting H2S mask
Function: Mitigation by removing persons
Procedural
Activity: tetection of H2S concentration and alarming when it exceeds limits
Personal protection measures - against excessive H2S and working with moist gas
Function: Individual protection against gas leak
Procedural
Activity: personal protection measures against H2S and moist with micro-organisms
A3: Gas Leakage in construction site (possbily intoxication)

Threats and preventive barriers on the left, accidents and mitigating barriers on the right. Edited from (RVO 2016)

STPA Based Approach for a Resilience Assessment at an Early Design Stage of a Cruise Ship

C. Bongermino[1] , P. Gualeni[1]
[1] University of Genoa (Italy)

corresponding author: Paola Gualeni – paola.gualeni@unige.it

ABSTRACT

Several definitions and approaches have been proposed to study resilience in different fields like materials, ecology, psychology and infrastructures. A general definition, applicable also to human-made or engineered systems, describes resilience as the ability to maintain capability in case of disruption.
Thanks to its systemic, top-down approach, STAMP (System-Theoretic Accident Model and Processes) has been already identified in literature as a very effective and "conductive" reference when reasoning about the possible need of resilience of a complex system. The STAMP-based tool named STPA (System Theoretic Process Analysis) establishes the following steps: identify system accidents, hazards; draw functional control structure; identify unsafe control actions (UCAs); identify accident scenarios; formulate decisions and recommendations. It focuses on what actually is in the hands of the system designer and operator i.e. the possibility to take action on hazards that can be eliminated or controlled.
In this paper an approach to design resilience into a cruise vessel will be proposed. An application case will be developed considering the specific hazard of dead ship condition i.e. of energy black-out on board. In case of navigation close to the shore and in heavy weather condition, this situation can rapidly evolve into a loss. The ship energy production and delivery system, both for the propulsion and for the hotel services, will be considered. Running the procedure up to the level of UCAs enables the identification of the possible disruptive events capable to degrade the operational performance of the system. Starting from this point, suggestions will be discussed for a selected UCA, able to prevent or mitigate it. A metric for ship resilience will be proposed as well with the aim to allow comparisons among different design solutions.

Keywords: STPA, Resilience; Dead-ship condition.

1. INTRODUCTION

The issue of power generation and delivery on board is very relevant for any kind of ship. It is well known for example that for both passenger and cargo vessels an emergency sources of electrical power shall be provided, for essential services under emergency conditions (IMO, 2014). Emergency generator and emergency switchboard of the ship should be located above the uppermost continuous deck, should have independent fuel supply and be capable of giving power for the period of 18 hours for the cargo ship and 36 hours for the passenger ship.

It should be capable of supplying simultaneously at least the following services, very basic:

- Emergency lightening (at the alleyway, stairways, and exits, muster and embarkation stations, machinery space, control room, main and emergency switchboard, firemen's outfits storage positions, steering gear room)
- Fire detecting and alarming system
- Internal communication equipment
- Daylight signalling lamp and ship's whistle
- Navigation equipment
- Radio installations, (VHF, MF, MF/HF)

- Watertight doors
- Fire pumps, emergency bilge pump

The dead-ship condition however is when the ship has just the emergency generator/s working in compliance with what above, but there is no power for ship propulsion and manoeuvring, neither for the hotel services. It is a very critical situation especially for cruise ships due to the significant number of human lives on board. In recent years characterized by the increment of cruise vessels size, the concept that the ship herself represents the best possible lifeboat has earned credit in the safety rules framework. Therefore, the need to guarantee a proper amount of energy, in addition to the traditional emergency generator support, has become evident. For passenger ships, the safety implications of power availability on board are so relevant that from June 2010 the Safe Return to Port standard (IMO, 2006) has been introduced. The regulation requires that passenger vessels with a length of 120 metres or more or with three or more main vertical zones is to be designed in such a way that in the event of a flood or fire emergency, passengers and crew can stay safely on board as the ship proceeds to port under her own power. Safe Rerturn to Port criteria defines a threshold where the ship's crew should be able to return to port without requiring passengers to evacuate. The power generation that is to be guaranteed onboard is meant not only for propulsion but also for the hotel services that can provide a sufficient level of vital comfort to passengers while the ship is on the way back to the safe port.

The overall functional requirements are intended to provide the following capabilities after an incident of fire or flooding:

— Ensure propulsion, steering, manoeuvring and navigational capabilities

— Ensure necessary service of the safety systems (fire safety and watertight integrity) in the remaining part of the ship that is not directly affected by the casualty

— Support safe areas for passenger and crew for the duration of the return to port voyage (e.g. water, sanitation, food, ventilation and light).

If the casualty extends beyond the defined threshold and the ship must be abandoned, the regulations require a limited number of systems to be remain available for 3 hours to facilitate an orderly abandonment.

The outcomes of Safe Return to Port in terms of design features of modern large passenger ships is an increased redundancy on board for propulsion, steering systems and electrical power delivery as well as new adapted architecture of safety or any other relevant systems.

Nevertheless it is well known that safety is not only a matter of redundancy and systems availability. In fact also for ships complying with the Safe Return to Port standards and the relevant implied redundancy, black-out is still an issue therefore worth to be investigated with a different perspective.

The dead-ship situation (i.e. the ship in black-out, the loss of energy for propulsion and minimum services vital for human beings) in fact is an emergency situation that can occur unexpectedly and in case of adverse weather condition and proximity to the shore it can rapidly evolve in ship loss.

In this paper an approach enabling the integration by design of resilience capability against black-out on board a cruise vessel will be proposed and discussed. The importance of a proper framework to model and discuss at an early design level interactions and integration among the energy system, the automation system and the human operators become evident during the application, evidencing also the importance of designing for operations.

The current evolution of safety paradigm (from safety-I to safety-II) defines safety as the ability to succeed under varying conditions. The understanding of everyday functioning is therefore a necessary prerequisite for the understanding of safety performance (Hollnagel et al. 2006; Hollnagel, 2016). In 'Safety II', humans are seen as a resource necessary for flexibility and resilience. But in an era where human error is considered the cause of the majority of maritime casualties, the view of humans as a safeguard and not a problem is one of the biggest challenge.

In this respect, a starting point for organizations interested in Safety II is to enhance their employees' resilience, as the ability to monitor things and handle situations (Hollnagel et al. 2015). At present and for the specific case of ships, these abilities have to be considered as the result of a virtuous integration with IT on board and in particular with the automation system (Rahimia & Madni, 2014).

Focusing on what goes right, rather than on what goes wrong, changes the definition of safety from ’avoiding that something goes wrong’ to ’ensuring that everything goes right’ (Hollnagel., 2014). The attitude implied in Safety-II is ensuring that things go right but the first step is to acknowledge the inevitability and necessity of performance variability, second to find ways to monitor it, and third to find ways to control it (Hollnagel, 2016).

To this aim it seems very helpful the use of STAMP technique and in particular of STPA and the Safety Control Structure appears to be effective to model and reason about the best ways to enforce and implement safety from the top to the bottom of the structure, by monitor and control.

2. STPA APPROACH FOR RESILIENCE ASSESSMENT

Systems-Theoretic Accident Model and Process (STAMP) is a top-down system-based accident model that focuses on enforcing constraints rather than preventing failures (Leveson, 2011). In this approach, safety is a dynamic control problem rather than a reliability problem. The system is described by a hierarchical Safety Control Structure in which each part of the system is identified and analysed with its relationship with the other parts of the system, underlining what they communicate and do. The main focus of this approach is to identify the safety constraints that are exerted from the Safety Control Structure, because events leading to a loss can occur only when safety constraints from a higher level in the Safety Control Structure are not enforced.

STAMP is used to come up with high-level list of hazards in which disruptive events could arise, considering each part of the system as a contributor to the ongoing development of the emergent behaviours properties.

System-Theoretic Process Analysis (STPA) is the hazard technique (Leveson, 2011) built upon the foundation provided by STAMP, that mainly consists of creating basic system engineering information, identifying unsafe control actions and identifying causal factors of unsafe actions.

The roadmap of STPA consists of: define the purpose of the analysis (identify losses and hazards, define system boundaries), model the control structure, identify the unsafe control actions, identify loss scenarios.

In this perspective STPA is applied as a very effective way to understand where and how performance variability might happen (Leveson at al. 2006). Therefore it can also suggest how to better handle situations. An interesting application of STPA to favour resilience integration is formulated in Beach et al. (2018) where a particular attention to the development of metrics is given in order to compare different resilience solutions.

The STPA approach can be assumed as a possible technique to spot the need of resilience when pursuing an emergent property like safety and to subsequently guide its implementation during the design process with a link to the operational life of a complex system and the involved human operators. The perspective is that safety represents the overall target and resilience is just an enabling mean or better “the ability of the system to monitor the changing risk profile and take timely action to prevent the likelihood of damage” (Madni & Jackson, 2008). As already mentioned, STPA focuses on behavioural safety constraints and it enables the analysis at the socio-organizational level. Therefore it can suggest the most appropriate level and “typology” of resilience that should be enforced to manage such aspects.

STPA outputs can be used in many different ways, among which:

- Drive the system architecture
- Create executable requirements
- Identify design recommendations
- Identify mitigations and safeguards needed
- Drive new design decisions (if STPA is used during development)

Therefore the four possible resilience modes i.e. avoiding, absorbing, adapting and recovering (Madni & Jackson, 2009) can be formulated and implemented in a logic of interactions and interfaces to manage (monitor and take timely actions) an hazard.

The integration of resilience can be performed by design methods grounded in experience. One of the most popular is the so called physical redundancy but since we are interested to overcome the traditional reliability, the functional redundancy should be considered at least as more promising for the purpose of this paper. Many other design methods can be mentioned and a comprehensive list is reported in Madni & Jackson (2009), useful for the last part of the application and significant because able to involve also crew members in the process of design for operations.

3. APPLICATION TO A LARGE CRUISE SHIP

A typical solution for cruise ships is the propulsion performed by electric engines, that are the main load for what concerns the electric power generation and delivery system on board. Nevertheless for large passenger ships, the sum of all the electric loads necessary for the ship operational profile (generally indicated as the “hotel loads”) is comparable to the electric load for propulsion. The shipboard power plant consists of electric generator units, for instance synchronous generators, that are usually coupled with turbines or diesel engines.

The power generated by the whole power plant is provided by different units and delivered to the main electrical panel (main switchboard) in medium voltage. For the ships with more than 3 MW installed on board, the Safety of Life at Sea Convention – SOLAS (IMO, 2014) requires that the main panel has to be splitted in at least two sections. The rated voltage usually used for the main panel are 3,3kV, 6,6 kV and 11kV. On board cruisers the rated voltage is usually 6,6 kV or 11 kV.

The power supply of the electrical users is ensured by the distribution grid, that is usually subdivided in primary and secondary grid. The first one is in medium voltage, the second one in low voltage. Low voltage is usually 690, 440, 230 and 120V. The primary grid supplies the loads that need high power, such as the propulsion engines, the thrusters and the air conditioning compressors. The low voltage switchboards power the loads that require limited voltage (i.e. 230 V, 120 V). Moreover, in order to supply high power required by some specific user groups, there are some substations to ensure a specific service, for instance the galley (440 V) or the engine room (690 V). The shipboard distribution grids can be structured in different ways, depending on the type of ship and the power installed on board, such as radial or ring grid.

An example of a typical power generation and distribution system of a large cruise ship is shown in figure 1 (Vie, 2014).

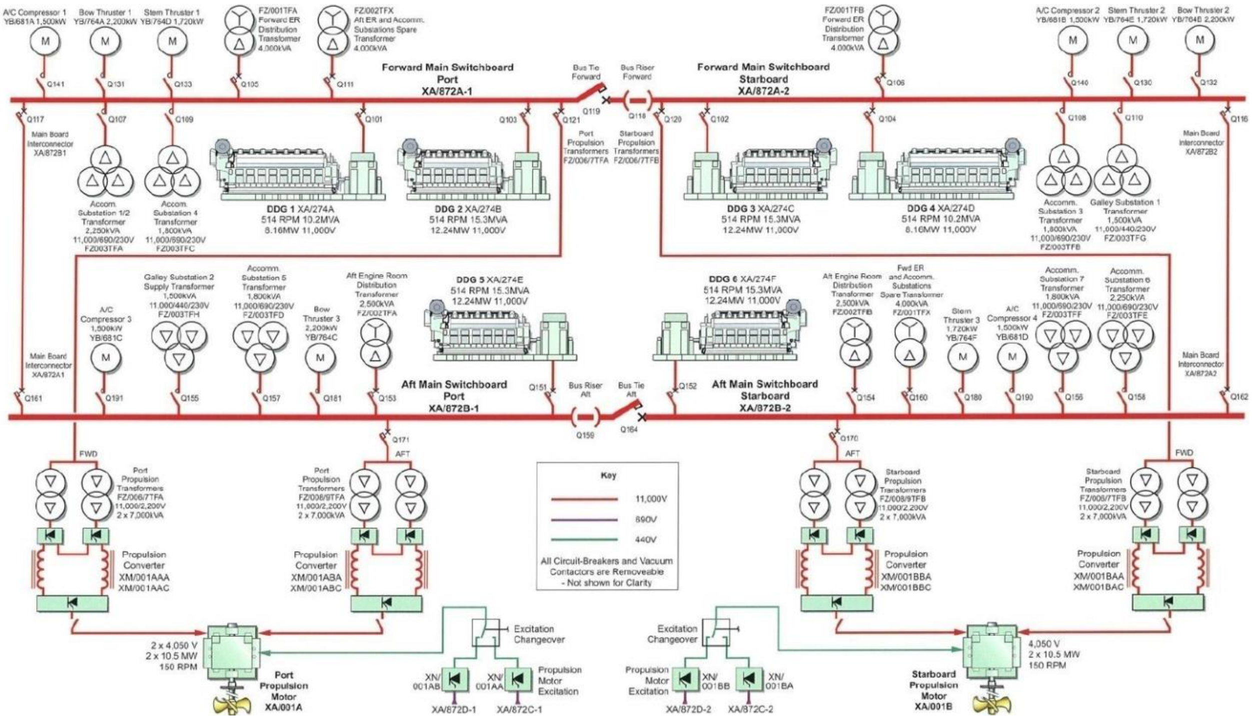

A/C Compressor 1 YB/681A 1,500kW
Bow Thruster 1 YB/764A 2,200kW
Stern Thruster 1 YB/764D 1,720kW
FZ/001TFA Forward ER Distribution Transformer 4,000kVA
FZ/002TFX Aft ER and Accomm. Substations Spare Transformer 4,000kVA
Forward Main Switchboard Port XA/872A-1
Bus Tie Forward
Bus Riser Forward
FZ/001TFB Forward ER Distribution Transformer 4,000kVA
Forward Main Switchboard Starboard XA/872A-2
A/C Compressor 2 YB/681B 1,500kW
Stern Thruster 2 YB/764E 1,720kW
Bow Thruster 2 YB/764B 2,200kW
Q141
Q131
Q133
Q105
Q111
Q119
Q118
Q106
Q140
Q130
Q132
Q117
Q107
Q109
Q101
Q103
Q121
Q120
Q102
Q104
Q108
Q110
Q116
Main Board Interconnector XA/872B1
Port Propulsion Transformers FZ/006/7TFA
Starboard Propulsion Transformers FZ/006/7TFB
Main Board Interconnector XA/872B2
Accom. Substation 1/2 Transformer 2,250kVA 11,000/690/230V FZ003TFA
Accom. Substation 4 Transformer 1,800kVA 11,000/690/230V FZ003TFC
DDG 1 XA/274A 514 RPM 10.2MVA 8.16MW 11,000V
DDG 2 XA/274B 514 RPM 15.3MVA 12.24MW 11,000V
DDG 3 XA/274C 514 RPM 15.3MVA 12.24MW 11,000V
DDG 4 XA/274D 514 RPM 10.2MVA 8.16MW 11,000V
Accomm. Substation 3 Transformer 1,800kVA 11,000/690/230V FZ/003TFB
Galley Substation 1 Transformer 1,500kVA 11,000/440/230V FZ/003TFG
A/C Compressor 3 1,500kW YB/681C
Galley Substation 2 Supply Transformer 1,500kVA 11,000/440/230V FZ/003TFH
Accomm. Substation 5 Transformer 1,800kVA 11,000/690/230V FZ/003TFD
Bow Thruster 3 2,200kW YB/764C
Aft Engine Room Distribution Transformer 2,500kVA FZ/002TFA
DDG 5 XA/274E 514 RPM 15.3MVA 12.24MW 11,000V
DDG 6 XA/274F 514 RPM 15.3MVA 12.24MW 11,000V
Aft Engine Room Distribution Transformer 2,500kVA FZ/002TFB
Fwd ER and Accomm. Substations Spare Transformer 4,000kVA FZ/001TFX
Stern Thruster 3 1,720kW YB/764F
A/C Compressor 4 1,500kW YB/681D
Accomm. Substation 7 Transformer 1,800kVA 11,000/690/230V FZ/003TFF
Accomm. Substation 6 Transformer 2,250kVA 11,000/690/230V FZ/003TFE
Main Board Interconnector XA/872A1
Main Board Interconnector XA/872A2
M
Q161
Q191
Q155
Q157
Q181
Q153
Aft Main Switchboard Port XA/872B-1
Q151
Bus Riser Aft
Bus Tie Aft
Q152
Aft Main Switchboard Starboard XA/872B-2
Q154
Q160
Q180
Q190
Q156
Q158
Q162
Q159
Q164
Q171
Q170
FWD
AFT
Port Propulsion Transformers FZ/006/7TFA 11,000/2,200V 2 x 7,000kVA
Port Propulsion Transformers FZ/008/9TFA 11,000/2,200V 2 x 7,000kVA
Starboard Propulsion Transformers FZ/008/9TFB 11,000/2,200V 2 x 7,000kVA
Starboard Propulsion Transformers FZ/006/7TFB 11,000/2,200V 2 x 7,000kVA
Key
11,000V
690V
440V
All Circuit-Breakers and Vacuum Contactors are Removeable - Not shown for Clarity
Propulsion Converter XM/001AAA XM/001AAC
Propulsion Converter XM/001ABA XM/001ABC
Propulsion Converter XM/001BBA XM/001BBC
Propulsion Converter XM/001BAA XM/001BAC
Excitation Changeover
2 x 4,050 V 2 x 10.5 MW 150 RPM
Port Propulsion Motor XA/001A
XN/001AB
XN/001AA
Propulsion Motor Excitation
XA/872D-1
XA/872C-1
XN/001BB
XN/001BA
XA/872D-2
XA/872C-2
4,050 V 2 x 10.5 MW 150 RPM
Starboard Propulsion Motor XA/001B

Figure 1: a typical layout of the power generation and distribution system for a cruise ship (Vie, 2014)

In this paper the attention will be focussed on a large cruise ship, during a very preliminary design phase when some reasoning about the black-out issue is very appropriate: it is both a safety issue (in case the ship propulsion is lost, especially in stormy weather close to the shore) and a commercial issue (in case the hotel service is lost with strong disappointment and discomfort for passengers). Usually, the starting point is a scheme like the one shown in figure 1. In an innovative perspective, the human factor, its integration with the automation system and the socio-organizational aspects as well, should be added into the discussion.

Following the STPA steps as mentioned in the previous paragraph, the hazards, the Safety Control Structure and the UCAs have been identified. Since the focus is on the black-out issue onboard, the identified hazards are the ones reported in table 1:

Table 1 the identified hazards for a focus on black-out on board

H 1	the ship propulsion is lost
H 2	the ship hotel services are lost

In figure 2 the Safety Control Structure sketched for the application is presented.

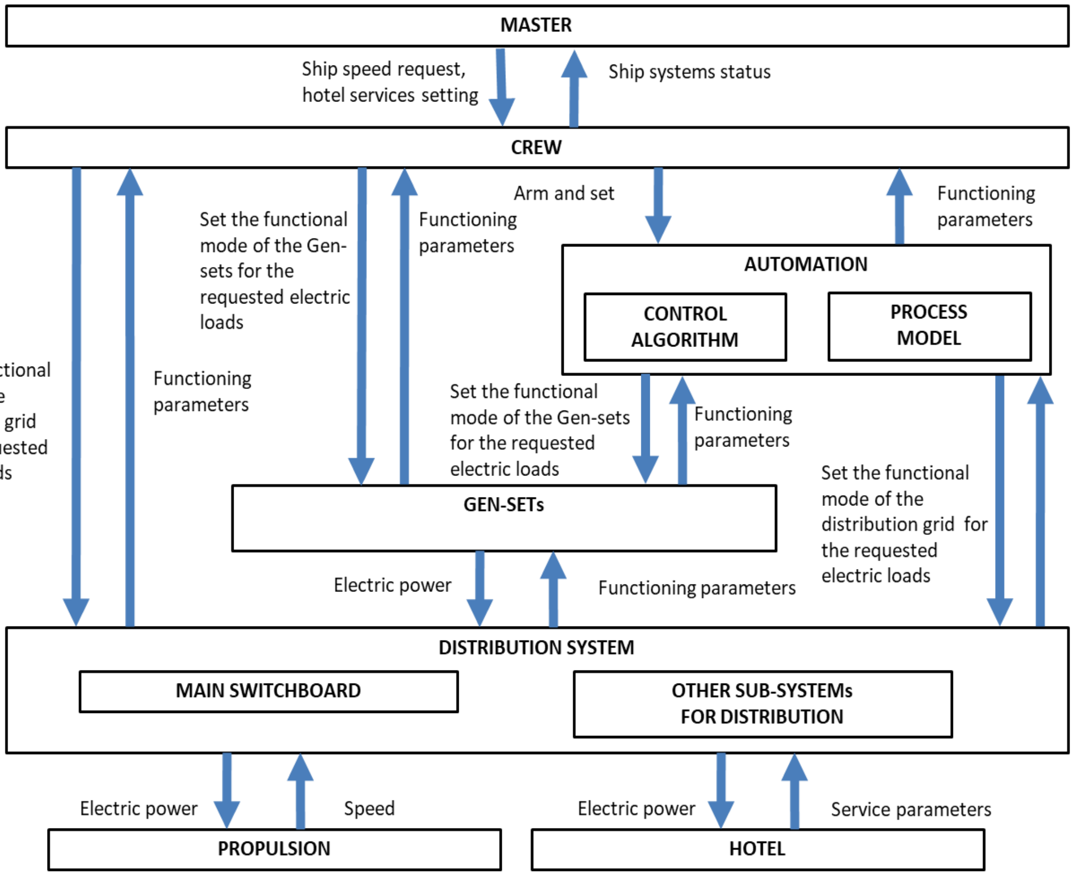

MASTER
Ship speed request, hotel services setting
Ship systems status
CREW
Set the functional mode of the Gen-sets for the requested electric loads
Functioning parameters
Arm and set
Functioning parameters
AUTOMATION
CONTROL ALGORITHM
PROCESS MODEL
Set the functional mode of the distribution grid for the requested electric loads
Functioning parameters
Set the functional mode of the Gen-sets for the requested electric loads
Functioning parameters
Set the functional mode of the distribution grid for the requested electric loads
GEN-SETs
Electric power
Functioning parameters
DISTRIBUTION SYSTEM
MAIN SWITCHBOARD
OTHER SUB-SYSTEMs FOR DISTRIBUTION
Electric power
Speed
Electric power
Service parameters
PROPULSION
HOTEL

Figure 2: the Safety Control Structure (GEN-SETs indicate the diesel generators, MSWB is the main switchboard and GRID is the general indication for distribution grid)

The diagram has been used then to derive the Unsafe Control Actions (UCAs). Such phase is very long and resource consuming, therefore for the purpose of this paper, only a sub domain of UCAs has been reported in table 2 (where GEN-SETs indicate the diesel generators, MSWB is the main switchboard and GRID is the general indication for distribution grid).

The considered UCAs are relevant only to the control and feed-back actions between the Automation System and the GEN-SETs. Power generators, in fact, they are assumed as one of the most important elements when analysing the black-out condition and their functioning is strongly dependent on the Automation System. This in turn means that the safety of the ship deriving from power delivery is strongly dependent on the proper Automation Systems actions.

A further selection will be made among the considered UCAs in order to create some examples to be finalized with proposal of resilience implementation. To this aim the attention has been focused only on one control action i.e. “set the functional mode of the Gen-sets for the requested electric loads" and UCAs have been formulated only for the class "not providing causes hazard" (UCAs from 1 to 9 in Table 2).

The further step, i.e. the definition of scenarios, means that two question are arising:

a) Why would Unsafe Control Actions occur?
b) Why would control actions be improperly executed or not executed, leading to hazards?

The definition of scenarios for all the UCAs mentioned in Table 2 would be too long and challenging for the purpose of this paper. Therefore only selected scenarios for UCA – 1 are formulated and reported, limiting the analysis to the area of the Safety Control Structure as evidenced in figure 3.

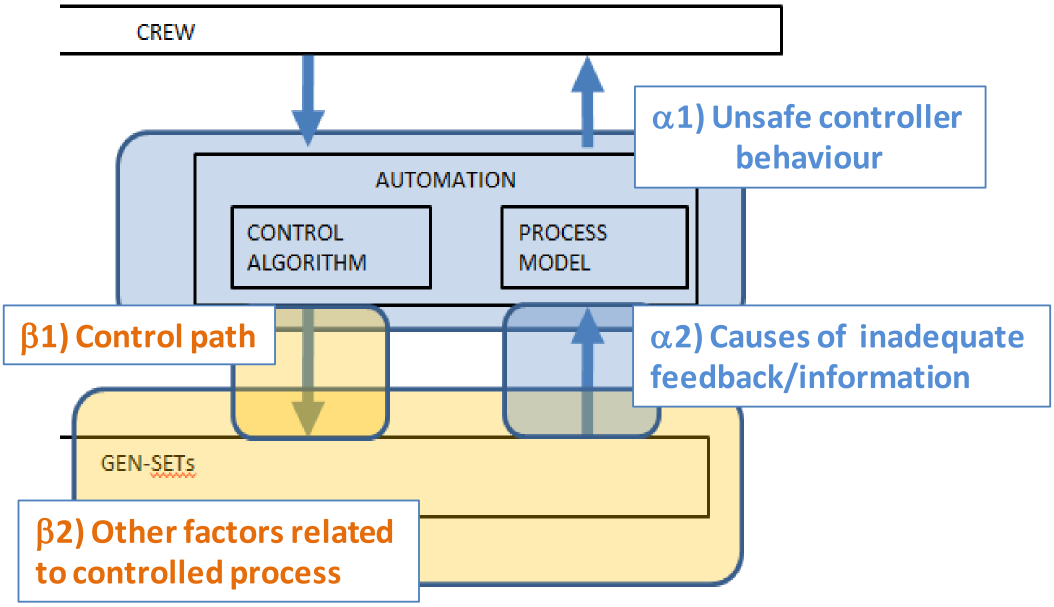

CREW
α1) Unsafe controller behaviour
AUTOMATION
CONTROL ALGORITHM
PROCESS MODEL
β1) Control path
α2) Causes of inadequate feedback/information
GEN-SETs
β2) Other factors related to controlled process

Figure 3: A specific focus on the Safety Control Structure in the perspective of scenarios definition

In the same figure 3 it is clarified that, thinking about the specific UCA, the two above mentioned questions a) and b) in turn requires to meditate on:

α1) Unsafe controller behavior
α2) Causes of inadequate feedback /information
β1) Control path
β2) Other factors related to controlled process

As described in Table 2, UCA-1 is: "AUTOMATION does not provide the functional mode of the Gen-sets for the requested electric loads during navigation [H 1, H 2]".

The identified scenarios are summarized in table 3.

For the purpose of this paper scenarios 2 and 3 are considered. It is worth mentioning that in some cases the sensors and the algorithm of the automation system can be challenged by the ship large motions when operating in extreme weather conditions.

With reference to them, a selection of design heuristics i.e. qualitative design methods grounded in experience are identified as a practical basis to provide resilience to the ship in operations for example in heavy seas.

Table 2 The list of Unsafe Control Actions (UCAs): a subset

Control Action	Not providing causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long
Set the functional mode of the Gen-sets for the requested electric loads	UCA-1 AUTOMATION does not provide the functional mode of the Gen-sets for the requested electric loads during navigation [H 1, H 2] UCA -2 AUTOMATION does not provide the functional mode of the Gen-sets for the requested electric loads during manoeuvring [H 1, H 2] UCA -3 AUTOMATION does not provide the functional mode of the Gen-sets for the requested electric loads in harbor [H 1, H 2]	NOT CONSIDERED FOR THE PURPOSE OF THIS PAPER	NOT CONSIDERED FOR THE PURPOSE OF THIS PAPER	NOT CONSIDERED FOR THE PURPOSE OF THIS PAPER

Table 3 The definition of scenarios for UCA -1

Scenario 1 for UCA – 1: the Gen-Sets controller (automation system) fails during navigation, causing an interruption of power delivery.
Scenario 2 for UCA – 1: the sensors report inadequately to the automation system that parameters are out of the safety range.
The Gen-Sets do not provide power to the ship in navigation because the automation system has ordered the shutdown: it detects that the engines are going to suffer a significant damage due to some functioning parameters out of safety range, due to inadequate sensor feedback.
Scenario 3 for UCA – 1: the specified control algorithm is flawed, so the automation system detects that the Gen-Sets parameters are out of the safety range.
The Gen-Sets do not provide power to the ship in navigation because the automation system has ordered the shutdown: it detects that the engines are going to suffer a significant damage due to some functioning parameters out of safety range, decided by a flawed control algorithm.
Scenario 4 for UCA – 1: the automation system sets the Gen-Sets parameters but this is not received by the system.
Scenario 5 for UCA – 1: the Gen-Sets suffers of a technical breakdown or malfunction.

From Madni & Jacknson (2009), among the fourteen design heuristics proposed by the authors, six could be defined as appropriate for the application:

- Functional redundancy: there should be alternative ways to perform a particular function that does not rely on the same physical systems.
- Human backup: humans should be able to back up automation when there is a context change that automation is not sensitive to and when there is sufficient time for human intervention.

- "Human in the loop": humans should be in the loop when there is a need for "rapid cognition" and creative option generation
- Intent awareness: system and humans should maintain a shared intent model to back up each other when called upon
- Learning/Adaptation: continually acquiring new knowledge from the environment to reconfigure, re-optimize, and grow
- Context spanning: system should be able to survive most likely and worst case scenarios, either natural or man-made.

Starting from these selection, In table 4 and 5 some proposals are made in order to implement resilience in relation with selected scenarios 2 and 3. The reason why such scenarios are selected is because they seem more appropriate to formulate resilience as the integration of operators, automation and design.

For scenario 2, functional redundancy, human backup and context spanning are selected as suitable design heuristic. For scenario 3 the learning/adaptation option has been preferred to the functional redundancy.

Table 4 Proposal for discussion of resilience implementation – UCA - 1 Scenario 2

	Functional redundancy	Human backup	Context spanning
Scenario 2	Subsidiary devices should provide information about parameters working point to assess whether they actually are outside the safety range	The operators should be able to make decisions independently from automation system and act accordingly. Possibly supported by subsidiary devices (see column on the left).	In the preliminary design all the possible operational scenarios have to be identified in order to define the operational domain of on board systems (to be assumed in the technical specifications, for example with reference to roll angle and/or list angle in heavy seas).

Table 5 Proposal for discussion of resilience implementation – UCA - 1 Scenario 3

	Learning/Adaptation	Human backup	Context spanning
Scenario 3	The control algorithm should be able to introduce in the logic of the procedure the awareness for example of stormy weather condition and in such case submit to human beings the decision about engine shutdown.	In specific cases like engine shut down the operators should be "consulted" by the automation system. Operators should receive the proper training for this.	In the preliminary design all the possible operational scenarios has to be identified in order to define the operational domain of on board systems (to be assumed in the technical specifications, for example with reference to roll angle and/or list angle in heavy seas).

From what above it appears how ship resilience is the result of an effective integration between operators and automation systems. This is a very important issue at present since automation is more and more exploited on board ships. The issue is even more important when automation has the total control on systems like GEN-SETs having a strong relation with safety: a stronger integration between human operators and the automation should be developed to drive a successful decision making for safety. The Safety Control Structure (the relevant part is reported in Figure 4) is very effective to put in evidence the hierarchy among them and the necessary control and feedback.

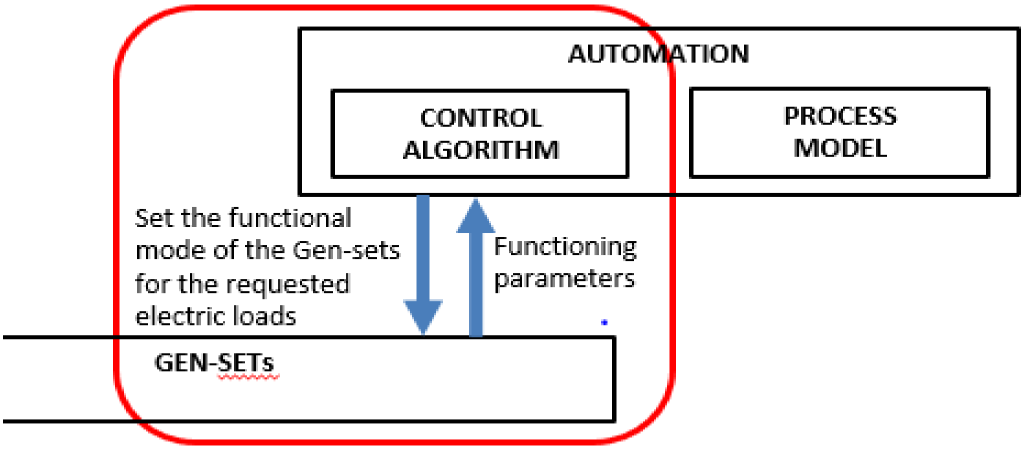

AUTOMATION
CONTROL ALGORITHM
PROCESS MODEL
Set the functional mode of the Gen-sets for the requested electric loads
Functioning parameters
GEN-SETs

Figure 4: A specific focus on the Safety Control Structure with GEN-Sets, Automation and Crew members connections

When mitigations and safeguards are identified it might be useful to investigate and compare different alternative solutions. In this perspective quantifiable metrics for resilience are in principle necessary.

Of course more than one indicator can be used and moreover a proper characterization also in terms of costs could help to better appreciate the cost/benefit ratio of the alternatives under investigation (Yodo & Wang, 2016).

With an approach based on STPA, which focuses on the possibility to take action on hazards that can be eliminated or controlled, it seems natural to define the metric hinged on the identified hazard.

For the proposed application, the identified hazard is the missing or insufficient power delivery. A possible indication for the purpose of this paper is to define as a quantitative indicator the percentage of available power with respect to the total power needed, as described below.

$$\varepsilon\ (t) = \frac{P\ (t)_{delivered}}{P\ (t)_{needed}}$$

It ranges from 0 to 1. When $\varepsilon = 1$ it means that the implemented resilience makes possible the complete delivery of the necessary power. The possibility to monitor with subsidiary devices the GEN-SETs could be implemented for all the units or just the number considered sufficient for avoiding the situation of total black-out. When $\varepsilon = 0.5$ it means that only one half of the needed power is available. Whether it is sufficiently safe or not is to be decided assuming criteria that set the minimum power necessary for propulsion (the weather condition should be considered in this case) and for hotel services.

The resilience by human backup is strongly linked with the provision of subsidiary information and training since the decision to interfere with the automation system should be based on the possibility to increase the situational awareness in terms of safety.

Finally it is worthwhile mentioning that this kind of metric is able to quantify the effect of resilience over a specific issue like electric power production and delivery. The assessment of an overall and more comprehensive ship resilience is, in principle, possible but very complex.

5. CONCLUSIONS

In the paper the possibility to apply STPA technique has been investigated in order to find out where and how resilience should be implemented on board relying on appropriate design heuristics.

An application case has been carried out with reference to a large cruise vessels. The specific issue of black-out on board has been selected and the hazards of propulsion and/or hotel services loss have been identified. Relying on the Safety Control Structure, a selection of Unsafe Control

Actions has been reported. One UCA has then been selected for the development of scenarios and the relevant need for resilience is spotted out.

STPA has enabled the visualization of the hierarchy among the ship energy system, the automation system and the crew members useful to discuss in a design stage the characteristics and the logic of the automation system (integration with crew members in decision making included), especially when some disruptive conditions like extreme ship motions can characterize the scenario and make things difficult for the automation system reliability.

The implementation of resilience has been proposed in terms of functional redundancy, learning/adaptation, human back up and context spanning. It has been put in evidence, in a design for operations perspectives, how the capability of a better integration between humans and the automation systems is envisaged in such a way that system should allow for human intervention needed without requiring humans to make unsubstantiated assumptions.

REFERENCES

Beach, P.M., Mills, R.F., Burfeind, B.C., Langhals, B.T., Mailloux, L.O., (2018) A STAMP-Based Approach to Developing Quantifiable Measures of Resilience, 16th Int'l Conf on Embedded Systems, Cyber-physical Systems, and Applications, Las Vegas

Hollnagel, E., Woods, D.P., Leveson, N. (2006) Resilience Engineering: Concepts and Precepts, Aldershot, pp 397.ISBN 0754646416

Hollnagel, E. (2014) Safety-I and Safety-II: The Past and Future of Safety Management CRC PressISBN 9781472423085

Hollnagel, E., Braithwaite, J., Wears, R. (2015) From Safety-I to Safety-II: A White Paper Technical Report · DOI: 10.13140/RG.2.1.4051.5282

Hollnagel, E. (2016) Resilience Engineering: A New Understanding of Safety, J Ergon Soc Korea 2016; 35(3): 185-191

IMO 2014 SOLAS consolidated edition

IMO 2006 Resolution MSC.216(82) Amendments to the International Convention for the Safety of Life at Sea, 1974, as amended

Leveson, N.G., Thomas, J.P. (2018) STPA Handbook

Leveson N.G. (2011) Engineering a safer world: Systems thinking applied to safety, MIT Press

Leveson, N., Dulac, N., Zipkin, D., Cutcher-Gershenfed, J. Carroll, J. Barrett, B. (2006) - Engineering Resilience into Safety-critical Systems - MIT – Boston - USA

Rahimia, M., Madni, A.M., (2014) Toward A Resilience Framework for Sustainable Engineered Systems, Procedia Computer Science 28, pp. 809 – 817

Madni, A.M., Jackson, S. (2009) Towards a Conceptual Framework for Resilience Engineering, IEEE Systems Journal, Special Issue in Resilience Engineering.

Vie, R. (2014) President’s Day Lecture - The Design and Construction of a Modern Cruise Vessel, Institute of Marine Engineering, Science and Technology, IMarEST London.

Yodo, N., Wang, P., (2016) Engineering Resilience Quantification and System Design Implications: a Literature Survey Journal of Mechanical Design Vol. 138

Border crossing point as a socio-technical system: applying STAMP and STPA to border security

Laura Salmela[1,*], Eetu Heikkilä[2], Sirra Toivonen[3] and Risto Tiusanen[4]
[1,3] Risk and Asset Management, Technical Research Centre of Finland VTT, Finland
[2,4] Machine Health, Technical Research Centre of Finland VTT, Finland

ABSTRACT

Digital transformation of European border management entails multiple benefits. Novel digital systems and cyber-physical infrastructures, such as biometrics-enabled automated border control and advanced analytics aim to equip border agencies with more effective tools against constantly evolving border security threats. Additionally, they provide new means to increase the agencies' performance in the context of growing volumes in trade and travel. Besides a myriad of opportunities for improved border management, the development, deployment and sophistication of these technologies compared to legacy systems bring about new vulnerabilities that may be hard to identify and manage with techniques used today.

This paper employs a systems-theoretic approach to address the security of border control systems. The focus is on border checks, which involve technologies used for ensuring and controlling that persons and the objects in their possession are authorised to enter or exit the EU area at external borders. The paper provides a preliminary review of current literature and discusses the basic tenets and main features of security analyses in this field by reflecting them against the STAMP model and the STPA technique for security analysis purposes.

The systems-theoretic approach is demonstrated in this paper by presenting the first phase of a coarse STPA-inspired security analysis at air borders with a particular interest on automated border control systems. Based on the analysis, STPA was found as a suitable approach for security analysis, as it supports assessment of the interactions between various stakeholders within the border control system. As a conclusion, we also provide insights for future research directions in cyber-physical border check systems and applications of systems-theoretic analysis methods in this particular field.

Keywords: border management; digitalisation; systems theory; STAMP; STPA-Sec.

1. INTRODUCTION

Since their first introduction in 2007 through the Portuguese RAPID programme (Frontex 2012), automated border control (ABC) systems have gradually become an established practice at Europe's external borders and a key component of smart, digitalised border control systems. Throughout the years, ABC systems have undergone significant development phases, and they have evolved "from biometric-enabled access-control-like systems operated by a set of known habituated users towards complex e-border systems operated by a flow of unknown non-habituated users" (Gorodnichy 2015, p. 59). When compared against the former systems, ABC systems now include a constantly growing number of non-biometric components, such as technologies for travel document verification, behavioural analysis of persons inside and within close proximity of the ABC gates and passenger risk assessment based on the use of various data sources. Total user volumes have also significantly increased, and the shift in user group

---

[*] Corresponding author: +358407498548 and laura.salmela@vtt.fi

characteristics from mostly experienced to mostly inexperienced has influenced the scope of human factors that can be adequately controlled during passenger processing in immigration control. (Gorodnichy 2015)

Deepening sophistication of ABC systems has had an impact on the ways in which their performance should be evaluated, also from the point of view of security. International standards designed for individual sub-components of the system, such as ICAO guidelines for machine-readable travel documents (e.g. Doc 9303) and ISO standards on biometrics (e.g. ISO/IEC 19794-5), have been considered too narrow to sufficiently account for “all system components and factors and their relationship with each other” (Gorodnichy 2015, p. 59). Traditional approaches to evaluate performance according to matching error rates may not provide sufficient results if a biometric system need to be assessed comprehensively (Thieme 2009). Current security analysis techniques focus on analysis of specific technologies or information security, but mostly lack the capability to perform comprehensive analysis regarding the complex, human-machine interaction at the actual border control point. Thus, system-theoretic approaches were found as a potential way to approach the problem, although their use is largely unexplored in this domain.

This paper is constructed as follows. Section 2 provides a short theoretical background by discussing the application of STAMP and STPA for security critical systems. Section 3 provides a brief description of the system in focus, while Section 4 presents a short literature review on what kind of approaches have previously been applied in analysing the security of/risks related to automated border control systems. Section 5 presents the preliminary results of a coarse STPA analysis, and Section 6 concludes the paper by reflecting the results against current development trends and present research.

2. STAMP AND STPA FOR SECURITY CRITICAL SYSTEMS

STAMP approach formulated by Leveson (2012) is a safety analysis approach based on systems theory. Contrasting with traditional component failure oriented approaches, STAMP defines safety as a control problem and models the system from the point of view of safety control. This provides efficient means for identifying hazards arising from the various interactions within the system. STAMP-associated hazard analysis method STPA provides a systematic process for identifying unsafe control actions within the system that may lead to losses. STAMP approach and STPA analysis have been mostly used in the context of safety, either to analyse existing systems or to support the investigation of accidents. Recently, the approach has seen increasing use also in the context of security-critical systems. This has spurred the development of some modifications and additions to the basic STPA, most notably the STPA-Sec (Young & Leveson, 2013) and STPA-SafeSec (Friedberg et al., 2017), both intended for safety and security co-analysis. Further expansions and improvement actions have been proposed by Schmittner et al. (2016) and Procter et al. (2017).

STPA-Sec provides a top-down approach for assessing a system from a security perspective. While STPA focuses on various unintended disruptions to a system, STPA-Sec extends the approach to account for intentional attacks towards a system. STPA-SafeSec provides a further process model for clearly taking into account both safety and security issues.

Existing STAMP-based studies on border control systems are not known to the authors. However, STPA and STPA-Sec have been recently applied in a number of cases dealing with various critical infrastructure related aspects. Some examples of applications relevant to the scope of this paper include the following:

- Williams (2015) has applied a STAMP-based approach on port security, arguing that, while preliminary in nature, the study suggests benefits in the paradigm shift from preventing failures towards enforcing security control actions.
- Beaumont & Wolthusen (2019) have successfully applied STPA-Sec in the context of critical infrastructure in the energy sector.
- The STPA methodology has further been elaborated and applied by Shapiro (2016) to account also for privacy-related issues, which is a key theme also in border security.

STPA and its extensions are fairly new methods in the research of security critical systems. Based on the studies referenced above, it is a promising approach for assessment of complex, socio-technical systems also from the security perspective. However, it should be noted that in the context of border security, a security breach does not necessarily result from an intentional activity, but can be also caused through various unintentional events, while still being considered as a security issue. In this paper, the general STPA approach is applied, as a distinction between security and safety related issues in this context is not seen relevant by the authors.

3. SYSTEM DEFINITION

By approaching the border check process from a general perspective, the core task of border authorities is "to determine whether persons are authorised to enter or leave the territory of a state, including checking their means of transportation and the objects in their possession and processing them accordingly" (European Commission 2010, p. 29). The main objective of an ABC system thus is to automate the authorisation control task. Border guards as human operators remain a very important part of ABC systems "as they make the ultimate decision about whether a person of interest has been identified" (Graves et al. 2011, p. 154).

The main functions of an ABC system are:

1) to authenticate the travel document of the passenger,
2) to verify the passenger's identity by connecting the passenger to the travel document,
3) to check the eligibility of the passenger to cross the border,
4) to permit or deny the passage of the passenger, and
5) to control the overall security of the process e.g. by providing alerts on tailgating attempts (Frontex 2015a).

The physical infrastructure of present ABC systems in operational use consist of two major sub-systems: **an e-gate(s)**, which works as a physical barrier and with which the passenger interacts; and **a control station(s)** from which a border guard(s) monitors the automated process. The main components of an e-gate are (a) electric doors, (b) passport reader, (c) biometric capture device(s), (d) system management hardware and software, and (e) monitors for passenger guidance. Besides physical barriers, the automated process may also include a separate kiosk, where the passenger completes certain parts of the process (e.g. enrols his/her live biometrics to the system) before crossing the border through the physical barrier. The control station is equipped with a user interface for monitoring and intervening the automated process.

ABC system boundaries may be drawn beyond the physical infrastructure. The definition of the system boundary for ABC much depends on the selected topology and physical design in a border crossing point[†]. For example, MacLeod and McLindin (2011) include within the ABC system boundary also the guidance of the passengers into automated procedures prior to actual border check formalities (e.g. in-flight instructional video, and signage at the airport). In addition, the manual border control counter(s) designed to handle exceptional situations and passengers, who do not succeed in using the e-gates though being eligible, are incorporated.

4. ENSURING SECURE BORDER CONTROL SYSTEMS

Evaluating the performance of biometric systems is highly important, because they can create failures, as none of the biometric modalities is free of errors (Gorodnichy 2009). In the generic guidelines on ABC implementation produced by the European Border and Coast Guard Agency Frontex, border agencies are advised to evaluate system risks particularly in the planning phase of a deployment process. Specifically, authorities should focus their assessment on the technical and operational requirements, which define the system's biometric matching performance and ensure secure flow of data. Software security on the other hand should be managed in cooperation with the technology vendors. In addition to this, the system's risk assessment should address user-related aspects, as the automation of control procedures changes the traditional work tasks of border guards and may create resistance within the organisations. Change management is suggested as an appropriate tool for this. (Frontex 2015a)

[†] A comprehensive description on current topologies used for ABC system across Europe can be found for example in the Best Practice Operational Guidelines for Automated Border Control (ABC) drafted by Frontex (Frontex 2015).

Once in operational use, the ABC system of a Member State may become a subject of an external EU-level evaluation and monitoring mechanism as specified in the Council Regulation 1053/2013. The objective of the regulation is to ensure that Member States meet the requirements established in the Schengen acquis (Council of the European Union 2013). The European Commission leads the evaluation system, but the Member States "influence the implementation of the evaluation mechanism as a whole" (Kaasik & Tong 2019, p. 14). The ABC system may also become evaluated through Frontex' vulnerability assessments (VAs) which aim to examine "the capacity and readiness of Member States to face upcoming challenges, including present and future threats at the external borders". The assessments focus on "the availability of the technical equipment, systems, capabilities, resources, resources, infrastructure, adequately skilled and trained staff". (European Parliament 2016a) The methodologies used for the Schengen evaluation and Frontex' vulnerability assessment are not open to public.

In research, several methods for the evaluation of ABC system security have been proposed. Individual international standards or standard families/frameworks, such as the ISO/IEC 27000 on information security management or ISO/IEC 31000 on risk assessment are considered to offer comprehensive tools for assessing system vulnerabilities (Heikkilä et al. 2017). Authors have also focused on particular aspects or sub-components of the system. With regards to data management, Schumacher (2017) proposes that a Data Protection Impact Assessment (DPIA) should be carried out in the design phase of the system. In contrast, Spreeuwers, Hendrikse & Gerritsen (2012) and Opitz & Kriechbaum-Zabini (2015) have evaluated the performance of the face recognition technologies, while Anand et al. (2016) examined multimodal biometric recognition (i.e face and fingerprint) and suggested means to enhance biometric data fusion. In their performance assessment methodology, also MacLeod and MacLindin (2011) address system security from a component perspective, namely focusing on the matching performance of the biometric algorithm. All these analyses, however, focus on specifics of the technologies applied within ABC systems, but do not provide a wider systemic perspective on the entire socio-technical system. Thus, new analysis methods need to be

5. PRELIMINARY STEPS OF STPA ON AUTOMATED BORDER CONTROL

Based on the EU regulation 2016/399 (Schengen Borders Code) governing cross-border traffic within the European Union and across its external borders (European Parliament 2016b), an indicative, exemplary list of high-level system losses and hazards may be derived as described in Tables 1 and 2. Identification of hazards is also based on previous literature and research work conducted in the FastPass project.

The Schengen Borders Code can be said to emphasize three aspects in particular: rights, security and respect of the rule of Community laws. The hazards may originate from risks internal to a Member State(s) (own nationals) or the border check process managed and implemented by the competent authority of a Member State. On the other hand, risks can be inflicted by actors external to Member States (e.g. terrorism, cross-border crime conducted by non-nationals or non-EU nationals). The first step of STPA constitutes definition of the analysis purpose, based on description of system losses (Table 1) and system-level hazards (Table 2).

Table 1. Description of system losses.

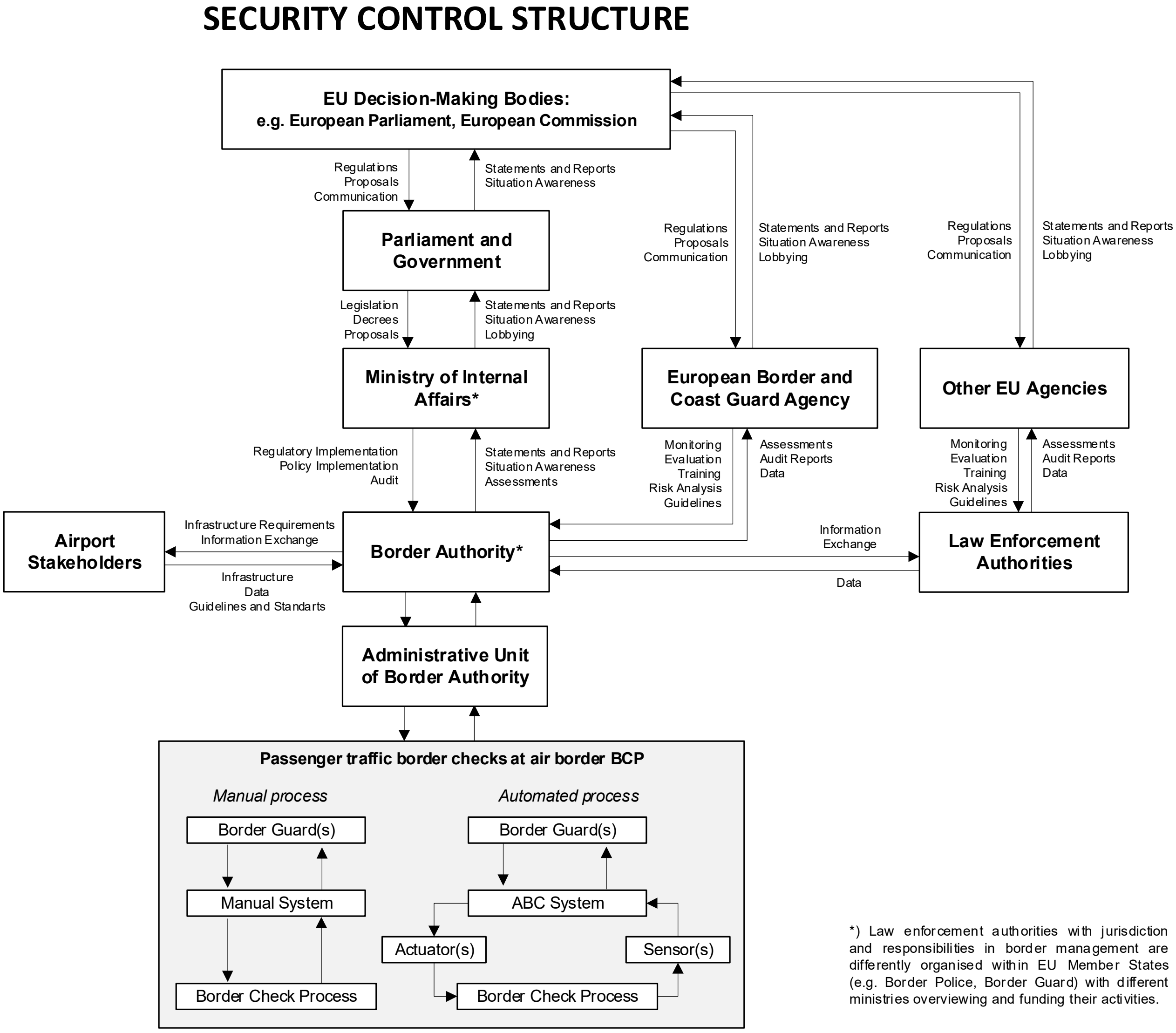
Losses	Description
L1	Violation of rights of free movement
L2	Violation of fundamental rights
L3	The internal security of any of the Member States is jeopardised
L4	External border control not carried out in accordance with EU regulations

Table 2. Description of system hazards resulting in high-level system losses.

Hazards	Related losses
H1 Traces of cross-border movement of persons enjoying the right of free movement are recorded	L1
H2 The performance of technologies is biased against different groups of people (e.g. biometric system)	L2

H3 Person of interest enters or exits the EU area undetected	L3
H4 Serious deficiencies in technical devices used to conduct border checks	L1, L2, L3, L4
H5 Border guards do not operate technical devices according to established guidelines	L1, L2, L3, L4

The security control structure for preventing the realisation of the identified losses and hazards is depicted in Figure 1. In the figure, both vertical and horizontal interactions between different system levels are presented in terms of control actions and feedback. Downward arrows and arrows leaving from the *Border Authority* indicate control actions, while upward arrows along with arrows leading to the *Border Authority* reflect feedback. The hierarchical order of stakeholders being positioned horizontally is not completely straightforward. For example, *Airport Stakeholders* include actors that operate at the international level and those who have activities only within a single airport. However, in order to keep the model simple, the stakeholders were grouped into larger units. In addition, other law enforcement authorities may pose requests (i.e. control actions) towards the Border Authority in connection with their specific mission tasks. Nevertheless, also in this case, this was discarded from the figure. Table 3 provides a generic, non-exhaustive description on the responsibilities of involved stakeholders ranging from local, regional and national level actors to international groups and organisations. The process owner for border checks both in manual and automated operational models is an administrative unit of a border authority that governs a particular border crossing point or border checks at a terminal facility (e.g. airport). The border guards who operate the border control system work under the administrative unit.



*) Law enforcement authorities with jurisdiction and responsibilities in border management are differently organised within EU Member States (e.g. Border Police, Border Guard) with different ministries overviewing and funding their activities.

Figure 1. Security control structure for automated border control.

As an illustrative example, the following ABC system sub-components can be defined as actuators and sensors according to the established method of STPA (e.g. Thomas 2018)[‡]. With respect to the actuators and sensors, one needs to bear in mind that the automated border check process is initiated by the traveler who places the travel document onto the document reader. As noted in ABC studies focusing on usability issues (e.g. Ylikauppila et al. 2014), this is a critical step potentially having a major impact on the success of the whole process or the total crossing time.

Actuators	***Sensors***
• Processing unit performing optical document checks, accessing and reading document data, verifying document data	• Document reader (incl. a radio frequency reader module)
• Processing units (1) initiating the capture of biometrics, (2) verifying capture biometrics	• Biometric capture device (camera, fingerprint reader)

Table 3. Description of stakeholder responsibilities in the security control structure.

Stakeholder	**Responsibilities**
EU Decision-Making Bodies	Proposes and establishes initiatives and policies and sets legislation and regulations on border security within the EU
European Border and Coast Guard Agency	Provides operational and technical assistance to MSs, monitors the situation and risk analysis at external borders, assists MSs in returns of nationals in non-EU countries, develops training programmes, cooperates with international organisations
Law Enforcement Authorities	Performs risk analysis in fields under jurisdiction (e.g. customs risk analysis on transported goods), exchanges data with border authorities on persons of interest
National Parliament	Sets legislation and regulations on border security at the national level
National Government	Proposes and establishes policies related to border security at the national level
Ministry of Internal Affairs	Overviews the implementation of national and EU policies and regulations, funds the operations of the border authority
Border Authority	Implements national and EU policies and regulations, directs and overviews border control operations at national level, develops, plans, procures and maintains equipment and systems for border control, directs human resource management at national level
Administrative Unit of Border Authority	Manages border check process at border crossing point(s), collects passenger information from carriers, performs risk analysis, conducts border checks on passengers, their means of transport and objects in their possession
Airport Stakeholders	Transmit passenger information to border control authorities e.g. Advanced Passenger Information (API) or Passenger Name Record (PNR) data (i.e. carriers); provide sector-specific guidelines, standards and other tools to control, manage and improve activities in aviation and within the airport environment and the travel industry in general (e.g. ICAO, IATA)

[‡] A comprehensive technical specification for ABC systems can be found e.g. in Frontex 2015.

5. DISCUSSION AND CONCLUSION

As noted by Kukula et al. (2010), the performance of biometric system is not dependent only on the matching algorithm but also on the user, the environment and the interrelationships between them. Current research together with evaluation methods derived from international standards offer good tools for addressing the risks at the sub-system or component level or the risks of specific aspects of the system, such as information security. However, as the sophistication of the ABC systems increases and interactions between different components become more complex, a more comprehensive approach is required to better identify and emphasize the real origin of system vulnerabilities and to propose specific countermeasures for their control.

As shown by the preliminary application of the STPA analysis technique in this paper, a systems-theoretic approach offers promising steps towards this direction. In particular, the security control structure for the automated border control system illustrates exhaustively the roles and responsibilities of different stakeholders. Also, by taking the analysis further, the STPA serves as an efficient technique for highlighting areas, which are and which are not in the direct control of the ABC system owner or the technology vendor. For example, an ABC system owner may establish a requirement to the vendor that a biometric system should not introduce gender or racial bias thus avoiding the potential violation in fundamental rights (see e.g. the London Policing Ethics Panel 2019 on the use of live facial recognition). In contrast, advanced fraud techniques, such as image morphing used in travel document application processes (see e.g. Raghavendra et al. 2017), constitute a major threat to ABC system security and EU border security. However, efficient countermeasures against such threats cannot be resolved within a single Member State but require interagency and cross-organisational cooperation at several levels of the security control structure.

The work presented in this paper also provides directions for future research in the area of applying systems-theoretic approaches on border security. For instance, comprehensive STPA studies should be performed on both manual and automated border control activities to provide comparisons of their strengths and weaknesses. Furthermore, as the work presented here is a first exploration of STAMP and STPA in a new domain area, broader applicability studies are also needed.

ACKNOWLEDGEMENTS

This work is based on research insights generated within the FastPass (the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No312583).

We thank the anonymous reviewer for insightful comments, which ensured the delivery of a stronger publication.

REFERENCES

Anand, A., Donida Labati, R., Genovese, A., Munoz, E., Piuri, V., Scotti, F., & Sforza, G. (2016). Enhancing the Performance of Multimodal Automated Border Control Systems. International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, 2016, pp. 1-5.

Beaumont P., & Wolthusen S. (2019). Micro-Grid Control Security Analysis: Analysis of Current and Emerging Vulnerabilities. In Gritzalis D., Theocharidou M., Stergiopoulos G. (eds) Critical Infrastructure Security and Resilience. Advanced Sciences and Technologies for Security Applications. Springer.

Council of the European Union. (2013). COUNCIL REGULATION (EU) No 1053/2013 of 7 October 2013 establishing an evaluation and monitoring mechanism to verify the application of the Schengen acquis and repealing the Decision of the Executive Committee of 16 September 1998 setting up a Standing Committee on the evaluation and implementation of Schengen

European Commission. (2010). Guidelines for Integrated Border Management in European Commission External Cooperation. EuropeAid Cooperation Office.

European Parliament. (2016a). REGULATION (EU) 2016/1624 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 September 2016 on the European Border and Coast Guard and amending Regulation (EU) 2016/399 of the European Parliament and of the Council and repealing Regulation (EC) No 863/2007 of the European Parliament and of the Council, Council Regulation (EC) No 2007/2004 and Council Decision 2005/267/EC.

European Parliament. (2016b). REGULATION (EU) 2016/399 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 9 March 2016 on a Union Code on the rules governing the movement of persons across borders (Schengen Borders Code).

Friedberg, I., McLaughlin, K., Smith, P., Laverty, D., & Sezer, S. (2017). STPA-SafeSec: Safety and security analysis for cyber-physical systems. Journal of Information Security and Applications, 34, 183–196. https://doi.org/10.1016/J.JISA.2016.05.008

Frontex. (2012). Best practice operational guidelines for automated border control (ABC) systems. Retrived from https://publications.europa.eu/en/publication-detail/-/publication/9bb99ddf-f4ec-4d87-b9de-669cabc14f2a

Frontex. (2015a). Best practice operational guidelines for automated border control (ABC) systems. doi: 10.2819/39041. Retrieved from https://frontex.europa.eu/assets/Publications/Research/Best_Practice_Operational_Guidelines_ABC.pdf

Frontex. (2015b). Best practice technical guidelines for automated border control (ABC) systems. doi:10.2819/071801. Retrieved from https://euagenda.eu/upload/publications/untitled-6349-ea.pdf

Gorodnichy, D. (2009). Evolution and evaluation of biometric systems. Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Application (CISDA 2009).

Gorodnichy, D. (2015). Analysis of Risks and Trends in Automated Border Control. Contract Report DRDC-RDDC-2016-C324. August 2015. Retrieved from http://cradpdf.drdc-rddc.gc.ca/PDFS/unc229/p803869_A1b.pdf

Graves, I., Butavicius, M., MacLeod, V., Heyer, R., Parsons, K., Kuester, N., McCormac, A., Jacques, P., & Johnson, R. (2011). The role of the human operator in image-based airport security technologies. In L.C. Jain, E. V. Aidman, C. Abeynayake (Eds.), Innovations in defence support systems - 2. SCI volume 338, (pp. 147-181). doi: 10.1007/978-3-642-17764-4

Heikkilä, A., Kojo, H., Toivonen, S., & Zehetbauer, S. (2017). High security solution. In S. Toivonen & H. Kojo (Eds.) Recommendations for future ABC installations - Best practices, (pp. 72-75). Retrieved from https://www.vtt.fi/inf/pdf/technology/2017/T303.pdf

Kaasik, J., & Tong, S. (2019). The Schengen Evaluation Mechanism: Exploring the views of experts in the field of police cooperation. European Law Enforcement Research Bulletin Nr. 18. (Winter 2019).

Kukula, E., Sutton, M., Elliott, S. (2010). The Human–Biometric-Sensor Interaction Evaluation Method: Biometric Performance and Usability Measurements. IEEE Transactions on Instrumentation and Measurement, 59, (pp. 1-8). doi: 10.1109/TIM.2009.2037878

Leveson, N. (2012). Engineering a safer world: Systems thinking applied to safety. MIT Press.

London Policing Ethics Panel. (2019). Final report on live facial recognition. May 2019. Retrieved from http://www.policingethicspanel.london/uploads/4/4/0/7/44076193/lfr_final_report_-_may_2019.pdf

MacLeod, V., & McLindin, B. (2011). Methodology for the evaluation of an international airport automated border control processing system. In L.C. Jain, E. V. Aidman, C. Abeynayake (Eds.), Innovations in defence support systems - 2. SCI volume 338, (pp. 115-145).

Opitz, A., & Kriechbaum-Zabini, A. (2015). Evaluation of face recognition technologies for identity verification in an eGate based on operational data of an airport. 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Karlsruhe, 2015, pp. 1-5.

Procter, S., Vasserman, E. Y., & Hatcliff, J. (2017). SAFE and secure: Deeply integrating security in a new hazard analysis. In ACM International Conference Proceeding Series (Vol. Part F1305). ACM.

Raghavendra, R., Raja, K., Venkatesh, S., & Busch, C. (2017) Face Morphing Versus Face Averaging: Vulnerability and Detection. IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, 2017, pp. 555-563.

Schumacher, G. (2017) Data protection impact assessment for ABC systems. In S. Toivonen & H. Kojo (Eds.) Recommendations for future ABC installations - Best practices, (pp. 37-40). Retrieved from https://www.vtt.fi/inf/pdf/technology/2017/T303.pdf

Schmittner, C., Ma, Z., & Puschner, P. (2016). Limitation and improvement of STPA-Sec for safety and security co-analysis. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9923, pp. 195–209. Springer.

Shapiro, S. (2016). Privacy Risk Analysis Based on System Control Structures: Adapting System-Theoretic Process Analysis for Privacy Engineering. Proceedings of the 2016 IEEE Security and Privacy Workshops, ser. SPW ’16. IEEE, 2016, pp. 17–24

Spreeuwers, L., Hendrikse, A., & Gerritsen, K. (2012). Evaluation of automatic face recognition for automatic border control on actual data recorded of travellers at Schiphol Airport. BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), Darmstadt, 2012, pp. 1-6.

Thieme, M., (2009). Performance Testing Methodology Standardization. In Li S.Z., Jain A. (Eds.), Encyclopedia of Biometrics. Springer, Boston, MA, (pp. 1069-1074).

Thomas, J. (2018). How to do a basic STPA. In N. Leveson & J. Thomas (Eds.), STPA Handbook, pp. 14-53.

Williams, D. (2015). Beyond a series of security nets: applying STAMP & STPA to port security. Journal of Transportation Security, 8(3), 139-157.

Ylikauppila, M., Toivonen, S., Kulju, M., & Jokela, M. (2014). Understanding the factors affecting UX and technology acceptance in the context of automated border controls. IEEE Joint Intelligence and Security Informatics Conference (JISIC), The Hague, 2014.

Young, W., & Leveson, N. (2013). Systems thinking for safety and security. Proceedings of the 29th Annual Computer Security Applications Conference ACSAC ’13, pp. 1–8. ACM Press.

Towards maritime traffic coordination in the era of intelligent ships: a systems theoretic study

Eetu Heikkilä[1,*], Mikael Wahlström[2] and Göran Granholm[2]
[1] Machine Health, VTT Technical Research Centre of Finland Ltd., Finland
[2] Human Factors, Virtual and Augmented Reality, VTT Technical Research Centre of Finland Ltd., Finland

ABSTRACT

Coordination of maritime traffic has developed over centuries with the main purpose of decreasing collisions and groundings of vessels. It has evolved from rudimentary measures, such as lighthouses, into an increasingly digitized setting with technologies like satellite positioning services and traffic coordination systems, such as the Automatic Identification System (AIS). In the future, increasingly intelligent shipping practices are expected to set further requirements not only for the ships themselves, but also for the coordination systems in maritime transport. Advanced and reliable coordination is especially seen as a key enabler for remote operated and autonomous ships.

The introduction of autonomous and unmanned smart ships is likely to be gradual, and coordination techniques of different technology levels are likely to co-exist in the maritime setting for an unforeseen period: there will be highly connected intelligent vessels and those applying very basic means of perception and communication. Based on the Systems-Theoretic Accident Model and Processes (STAMP) approach, as well as the related STPA hazard analysis methodology, this paper presents a control structure of maritime traffic coordination as it is now, and provides an overview of STPA hazard analysis performed on the system. It also discusses changes foreseen in the structure due to changing means of coordination in the future, providing basis for better understanding of the risks and opportunities. Additionally, the paper provides insights on applicability of STPA on a new application area of maritime traffic coordination.

Keywords: Maritime traffic coordination, technology levels, STAMP, STPA

1. INTRODUCTION

Maritime traffic coordination is needed to avoid collisions and groundings of ships, as well as to ensure efficient flow of maritime traffic. The methods used for maritime traffic coordination have substantially evolved in the past decades, with the introduction of technologies such as satellite navigation systems and digital tools for route planning and optimization. As a notable ongoing development, the International Maritime Organization (IMO) is working on standardization of electronic navigation in the e-Navigation effort (IMO, 2018). In the future, maritime traffic is expected to become increasingly autonomous, including also new means for maritime traffic coordination. Furthermore, meeting new energy efficiency targets requires new ways of managing marine traffic (Porathe et al., 2014). This is likely to result in situations where different vessels employing various levels of coordination co-exist at the same time.

To form an understanding of the various technology levels already present in the maritime setting and expected in the future, Wahlström et al. (2019) have proposed a categorization of Marine traffic coordination technology levels (Table 1). The categorization consists of six levels, each representing an increased level of coordination over the previous one. In Table 1, some examples are provided of each level. Currently, maritime traffic utilizes mostly coordination

---

[*] Corresponding author: +358408495790, eetu.heikkila@vtt.fi

methods ranging from levels 0 to 3, with developments ongoing to at least partially reach coordination level 4.

Table 1. A proposed categorization for Marine traffic coordination technology levels (adapted from Wahlström et al., 2019)

Marine traffic coordination technology level	Examples of coordination means
Level 0 – "I can see you": non-technological coordination only	Visual assessments, verbal communication, piloting activities
Level 1 – "I can see your flags": shared rules and passive communication mediums	Flags, lights, right-of-way rules, sea routes, lighthouses
Level 2 – "My radar sees you": signal-based detection, localisation and communication systems	Radar, radio communication, transponders
Level 3 – "I can see data about you": digitally enhanced coordination (sharing digitally stored data)	Automatic Identification System (AIS), waypoint and route-sharing using transponders or radio communication, Vessel Traffic Service (VTS), Electronic Chart Display and Information System (ECDIS)
Level 4 – "My robot sees you": smart coordination (localized/fleet/ship-level coordination based on machine learning and other predictive technologies)	On-board object detection systems, fleet management of intelligent ships
Level 5 – "My robot sees what your robot sees": internet of intelligent ships (global coordination based on machine learning, using shared data)	Artificial Intelligence (AI) optimized route generation, shared situational awareness, internet of intelligent ships and relevant land-based entities

2. METHODS

The aim of this paper is to provide views on potential changes in maritime traffic coordination due to introduction of technologies especially the highest levels, and especially due to the introduction of autonomous shipping. While details of the future direction of maritime traffic coordination and its policies are uncertain, it is certain that future maritime coordination constitutes a complex, socio-technical challenge. To understand the development of coordination in such context, a systemic approach is seen beneficial to identify the various interactions in maritime traffic coordination, as well as to provide a basis for further analyses regarding the risks of new coordination scenarios. Thus, to illustrate the coordination interactions, a systems-theory based approach was selected.

Specifically, the work performed in this paper follows the Systems-Theoretic Accident Model and Processes (STAMP) approach introduced by Leveson (2011). In STAMP, the analysed system is represented as a control structure, describing the various feedbacks within the system. In the maritime setting, STAMP has been utilized in various cases, such as in accident investigation (e.g. Kim et al., 2016), as well as in higher-level studies of maritime safety management (Valdez Banda & Goerlandt, 2018), (Aps et al., 2017). In this study, STAMP-associated STPA hazard analysis methodology (Leveson & Thomas, 2018) was applied on the maritime traffic coordination system, and based on this application, preliminary conclusions, as well as pointers for future research are presented.

3. APPLYING STPA ON MARITIME TRAFFIC COORDINATION

The basic process of performing STPA analysis has been described by Leveson & Thomas (2018). The process consists of four steps, starting from definition of the purpose of the analysis. In this case, the analysis background also has two underlying perspectives that should be introduced. Firstly, the aim is to provide insights for the maritime industry to form a better understanding of the future maritime coordination technologies, their interactions and related risks in the era of autonomous and intelligent ships, and especially to be better prepared for potential risks scenarios that may arise. Secondly, the study also serves as an exploration of STAMP and STPA in a new application area. So far, STAMP approach and STPA hazard analysis methodology have mostly been utilized in analysis of existing technologies or investigation of accidents that already occurred. In this paper, the approach is not only applied to analyse the current system, but also to support making predictions towards potential future scenarios with a number of uncertain elements. Thus, the work presented here should not be considered as complete hazard analysis, but rather as a means to support in building a view of current and future maritime coordination.

In the first step of STPA, the purpose of the analysis is defined based on identifying losses, system-level hazards, as well as system-level constraints. The losses considered here were defined as follows:

L-1: Loss of life or injury of people
L-2: Loss of or damage of vessel
L-3: Environmental loss

The first step also includes definition of the system boundary, i.e. what is to be included in the analysis. Here, a definition was made that only commercial vessels, and coordination measures relevant to them, are included in the consideration. Next, the identification of system-level hazards was performed, and the following system-hazards were identified:

H-1: Vessel operates in shallow waters or close to ground [L-1, L-2, L-3]
H-2: Vessels pass each other without sufficient passing distance [L-1, L-2, L-3]

To conclude the first step of STPA, system level constraints were identified as follows:

SC-1: Vessel must operate in marked fairways [H-1]
SC-2: Vessel must maintain safe passing distance [H-2]

The second step of STPA focuses on modelling the control system. In Figure 1, a STAMP control structure of maritime traffic coordination is presented. The model is based on a generalized view of the current situation with the examined vessel deploying typical coordination measures of marine traffic coordination technologies up to level 3 (see Table 1). As described in the previous step, the analysis is limited on commercial ships. It should also be noted, however, that all the coordination measures presented in Figure 1 may not be applicable to all commercial vessel types, and neither in all regions.

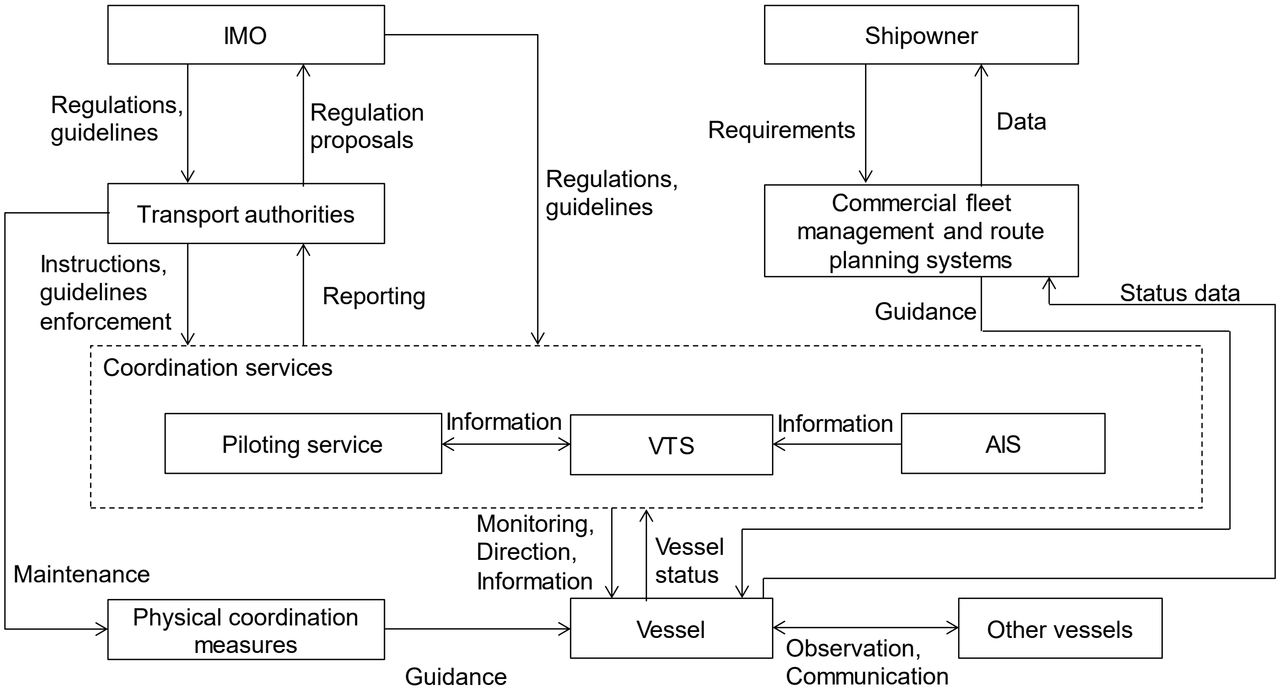


Figure 1. STAMP control structure describing current maritime traffic coordination, consisting of coordination means between levels 0 and 3 as described in Table 1.

The third step of STPA is to analyse the control structure to identify unsafe control actions (UCAs). To limit the scope of this paper, the study was focused on the control actions directly connected with the vessel. That is, the legislative and authority aspects pictured in the control structure were not studied in detail in this context. The identified UCAs are presented in Table 2.

Table 2. Identified unsafe control actions.

Control action	**Not providing causes hazard**	**Providing causes hazard**	**Too early, too late, out of order**	**Stopped too soon, applied too long**
Direction	UCA-1: Coordination services do not provide directions when Vessel is close to ground [H-1] UCA-2: Coordination services do not provide directions when Vessel is violating safe passing margins [H-2]	UCA-3: Coordination is provided based on incorrect vessel status information [H-1, H-2]	UCA-4: Coordination services provide directions too late for the vessel to react [H-1, H-2]	
Guidance (fleet management system)		UCA-5: Commercial fleet management system provides guidance that contradicts with coordination services [H-1, H-2]		
Guidance (physical measures)	UCA-6: Physical coordination measures are not in place correctly [H-1]			
Communication (other vessels)	UCA-7: Other vessels do not communicate their intentions when vessels are passing each other [H-2]	UCA-7: Communication by other vessels is incorrect and leads to misunder-standings when vessels are passing each other [H-2]		

The fourth and final step of STPA focuses on identifying loss scenarios. Based on the UCAs, a vast number of scenarios can be formulated. For example, UCA-1 above can be related to various scenarios, which may be related e.g. to issues within providing the coordination services, or incorrect feedback from the vessel. Similarly, a number of different scenarios can be identified for each of the UCAs. An exhaustive scenario listing is not provided here as it falls outside the scope of this paper.

4. DISCUSSION AND CONCLUSIONS

Based on the control structure presented in Figure 1, as well as the brief STPA analysis, it becomes apparent that currently maritime traffic coordination is a highly distributed system, as has also been argued e.g. by van Westrenen and Praetorius (2014). It consists of a number of separate means, like coordination services, various rules, communication practices, and other means often maintained by separate entities. The control structure also displays areas where feedback is missing, or where similar (potentially contradicting) control actions are provided by different entities. Thus, the focus is set on the responsibility of the vessel and its master, leaving the ship to optimize its own state, and emphasizing practices of good seamanship – a factor especially challenging to define and implement in autonomous shipping (Jalonen et al., 2018).

The continuing digital transformation in shipping, as well as the developments towards autonomous shipping, are likely to cause changes in the maritime traffic control structure. This is likely to introduce new risks, but also chances to improve the coordination of maritime traffic, which is currently distributed and reliant on the actions of individual vessels. STPA analysis was found promising in terms of identifying underlying issues in coordination systems. When considering a future situation where higher coordination levels are introduced, we can identify a number of potential changes based on literature findings and the control structure and STPA analysis above. The expected changes include:

- The role of commercial fleet management becomes increasingly important as ships become more autonomous. It is unclear how this will interact with the current services such as the VTS. Some sort of a centralized control system combining a number of actors that currently operate separately, may be beneficial or even necessary for autonomous shipping to develop. One example of such coordination is the proposed Sea Traffic Management approach by Lind et al. (2016).
- A situation where all vessels employ high-level coordination measures is unlikely in the near future. Instead, it is likely that vessels at coordination levels 4 and 5 still need to interact with vessels with lower level coordination measures. Thus, the ability to communicate in a way that can be interpreted by vessels at lower coordination levels will be necessary. In future research, separate STPA analyses of these different scenarios can enlighten such situations.
- Increasingly advanced on-board situational awareness systems can provide shared data for coordination, and may provide changes to the feedback loops in the control structure.
- Optimized coordination may also involve other actors that are not directly linked to the actual vessel operation but should be added to the control structure. For example, this can mean optimization of container shipping considering the entire logistics chain, including e.g. port and landside operations.
- In addition to changes in the vessels, some of the coordination measures (such as piloting service) may also be subject to substantial changes. Separated areas with differing coordination means are also possible.

Based on the above findings, a number of relevant themes for future research can be identified. Firstly, as the STPA analysis performed here was limited, it will be beneficial to expand the analysis to cover further aspects (some of which are already modeled in the control system in Figure 1) and to identify further unsafe control actions that are possible in the current coordination activities, potentially providing means for improving the safety of maritime transport. Additionally, a number of detailed scenarios, depicting future alternatives of various coordination strategies shall be studied to provide suggestions for further developing the coordination systems. Finally, the application of STAMP and STPA on coordination systems should continue to provide more rigorous analysis of its suitability in this type of problems.

REFERENCES

Aps, R., Fetissov, M., Goerlandt, F., Kujala, P., & Piel, A. (2017). Systems-Theoretic Process Analysis of Maritime Traffic Safety Management in the Gulf of Finland (Baltic Sea). *Procedia Engineering* (Vol. 179, pp. 2–12).

IMO. (2018). E-Navigation Strategy Implementation Plan – Update 1. International Maritime Organization.

Jalonen, R., Heikkilä, E., & Wahlström, M. (2018). Do We Know Enough About the Concept of Unmanned Ship? In P. K., & L. Lu (Eds.), Marine Design XIII: Proceedings of the 13th International Marine Design Conference (IMDC 2018) (Vol. 2, pp. 861-869). CRC Press.

Kim, T., Nazir, S., & Øvergård, K. I. (2016). A STAMP-based causal analysis of the Korean Sewol ferry accident. Safety Science, 83, 93–101.

Leveson, N. (2011). Engineering a safer world: systems thinking applied to safety. MIT Press, Cambridge, Mass

Leveson, N., & Thomas, J. (2018). STPA Handbook. Available: https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf

Lind, M., Hägg, M., Siwe, U., & Haraldson, S. (2016). Sea Traffic Management - Beneficial for all Maritime Stakeholders. Transportation Research Procedia (Vol. 14, pp. 183–192).

Porathe, T., De Vries, L., & Prison, J. (2014). Ship voyage plan coordination in the MONALISA project: user tests of a prototype ship traffic management system. Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference.

Valdez Banda, O. A., & Goerlandt, F. (2018). A STAMP-based approach for designing maritime safety management systems. Safety Science, 109, 109–129.

Wahlström, M., Heikkilä, E., & Granholm, G. (2019). Technology levels for maritime traffic coordination: towards the internet of intelligent ships. Abstract preprint available online: https://www.researchgate.net/publication/332073723_Technology_levels_for_maritime_traffic_coordination_towards_the_internet_of_intelligent_ships

van Westrenen, F., & Praetorius, G. (2014). Maritime traffic management: a need for central coordination? Cognition, Technology & Work, 16(1), 59–70.

Exploring the Modeling of Attack Strategies for STPA

Abdullah Altawairqi and Manuel Maarek
Heriot-Watt University (Edinburgh, Scotland, UK)

ABSTRACT

System analysis for security and for safety are both focused on identifying potential accidents and attacks, to implement prevention strategies. Security system analysis aims to counter intentional acts that could make the system vulnerable. Systems-Theoretic Process Analysis (STPA) is a holistic approach to system safety analysis. In this paper, we explore the possibility to combine STPA analysis with Attack-Defence Trees (ADTrees) modeling to strengthen a system security analysis. We also discuss how the identification of the intentions and capabilities of the attackers could focus the priorities of the analysis and reduce its scope. We suggest an approach on how to combine ADTrees' attack modelling and STPA to elicit unsecure control actions. To illustrate this approach, we apply it on a case study.
.

Keywords: STPA; attack defence tree; attack modeling; security

1. INTRODUCTION

System security analysis and safety analysis have in common the identification of the circumstances that could threaten the functions of the system or its integrity. While safety is concerned with avoiding accidents, the analysis of the security system aims to prevent the system from suffering from intentional acts. In security analysis, this modeling of the intention of an attacker is carried out with the identification of security targets and, in combination with the mapping of the attack surface, helps to specify the defence of the system to be built.

Systems-Theoretic Process Analysis (STPA) is a hazard analysis technique based on Systems-Theoretic Accident Model and Processes (STAMP) which structures the system analysis around the notion of control loops. Understanding how a control loop could malfunction or could fail to achieve its goal lead to identifying the system's safety constraints. STAMP integrates the notion of causal factor to elicit these safety properties. Such modeling is intrinsically focused on the point of view of the system while a security analysis should include the attacker's intention and capabilities. Because of its safety focus, STPA does not capture the attacker's intention in the analysis. Gleaning such intentions could lead to the identification of system vulnerabilities that are more likely to be used in an attack scenario. As the system's control loops are the core of STPA analysis, we propose to integrate their modeling of the system in the attack strategy modeling. Attack trees are an example of attack strategy modeling. Attack trees are a graphical representation for modeling and analysing potential attack strategies. They could be extended to consider defensive patterns in Attack-Defence Trees (ADTrees).

In this paper, we explore the possibility of combining STPA analysis and ADTrees modeling to strengthen a system security analysis. STPA is a top-down approach to identify unsafe control actions from the control structure. Each element of the STPA control loops of a system could be the direct or indirect target of an attack. Deriving ADTrees is in itself a top-down analysis so we suggest guiding its refinement process with steps to make explicit the way an attack impacts a control loop. A bottom-up approach to security analysis starts by considering the system's attack surface to evaluate how potential vulnerabilities could be exploited. We propose to integrate this attack surface perspective in our approach to combine ADTrees and STPA. To complement the modeling of attack intentions, we suggest to include attack profiles in our ADTree modeling to describe the potential attacker in terms of its skills and motivation. Associating attack profiles with attack scenarios help to narrow the scope of an analysis.

The structure of the paper is as follows: Section 2 gives background and related work on safety and security analysis approaches, Section 3 proposes an approach to integrate ADTrees with STPA, Section 4 applies the proposed approach to a case study of the steel plant, and Section 5 concludes.

2. BACKGROUND AND RELATED WORK

In this section we introduce the main background of our work, ADTrees and STPA. We then present some related works.

Attack trees (Schneier, 1999) are a graphical representation of the potential scenarios of an attack as a tree of potential attack strategies. Attack trees aim to provide a way of thinking about the system exposure to attacks. The root node of an attack tree is the goal of the attacker. The relationship between a parent node and its children nodes is following a logical structure called variance. Children nodes are either considered as conjunctions (AND) of actions that lead to the parent node state, or a disjunction (OR) of actions all resulting in the same parent node state. Attack trees have many flavours. Jhawar, Kordy, Mauw, Radomirović, & Trujillo-Rasua (2015) introduced the sequential conjunctive operator (SAND) that enforces an order in which the actions are to be conducted in the attack. Kordy, Mauw, Radomirović, & Schweitzer (2014) extended attack trees by considering defensive patterns in the so-called Attack Defence Trees (ADTrees). According to Kordy et al. (2014), original attack trees do not address the interactions between attacks scenario and the defences of the system. ADTrees contain defensive nodes called countermeasures. Those nodes could appear at any level on the tree and follow the logical structure AND and OR. Defensive nodes are system actions that are to prevent attack steps (Kordy et al., 2014). In ADTrees, a defensive node drawn as child or an attack node indicates that the attack is prevented by the defence.

STPA is safety analysis approach based on STAMP. STPA's approach focuses on accidental causes and safety constraints. STPA identifies the root of accidents which are hazardous scenarios to define safety constraints that need to be fulfilled by the system to prevent these accidents. It is top-down process to identify failure states of the system by analysing the controls of the system and how they can fail. This analysis leads to stating safety constraints the system must fulfil (Leveson & Thomas, 2013). STPA has four basic analysis steps. First, to define the purpose of the safety assessment, system losses and system hazards. Second, to identify the control actions of the system's control model. Third, to establish the safety constraints and requirements from the identified unsafe control actions. Fourth, to identify causal scenarios. While STPA guides the analysis in identifying causal scenarios leading to failed control loop, it does not provide guidance for the identification of intentional causal scenario based. A security causal scenario is characterised for instance by the attacker's intention, the attacker's capabilities, the system's surface of attacks.

A number of ongoing researches are proposing to extend safety system analysis for security. We discuss some of these works as they relate to the approach we are discussing in this paper.

STPA-Sec (Young & Leveson, 2013) aims at providing a solution to this security modeling need with a semi integrated approach between safety and security. It follows the STPA top-down approach but focuses on identifying losses and vulnerable states in order to strengthen the security of a system. STPA-Sec has the same basic process of STPA where vulnerabilities replace hazards. Even though STPA-Sec is an analysis approach for safety and security, it does not distinguish between intentional causal scenarios that are central to security analysis.

The Failure Mode Vulnerabilities and Effect analysis (FMVEA) is a step by step approach for investigating vulnerabilities-based failure mode and the potential effects these weaknesses could have in terms of decreased readability and availability of the system (Schmittner et al., 2014). It is an extension from The Failure Mode and Effect analysis (FMEA) used in safety to document the analysis of the impact of a component failure on the overall system. FMEVA proposes to include vulnerabilities and attack models to identify potential attack vectors of concern for the system. FMVEA uses cause effect chains into vulnerabilities, threat agents, threat modes, threat effects and attack probabilities in its modeling of attacks.

STPA-SafeSec (Friedberg, McLaughlin, Smith, Laverty, & Sezer, 2016) is a fully integrated approach between combining safety analysis and security analysis. The authors explain that their approach goes beyond STPA-Sec as it provides guidance to evaluate the safety impact the constraints derived from the security analysis could have. STPA-SafeSec extends the core of STPA's approach by considering security causal factors on integrity and availability. It claims to overcome limitations in STPA-Sec's approach by adding physical components layer into the control loop analysis to model the surface of attack and its link with the core safety features of the system. It also advocates for mapping security and safety constraints to the control layer in order to mitigate potential safety and security conflicts.

S-cube were introduced as a joint safety and security analysis model for industrial control system (Kriaa, Bouissou, & Laarouchi, 2015). S-cube is enabling formal modeling for system architecture and automates the generation of attack and failure scenarios. The automation results are depending on assigned hypothesis.

3. INTEGRATING ATTACK MODELING AND SYSTEM ANALYSIS

In this section, we will explain the proposed approach of the attack model for system security analysis. In section 3.1 we will present a brief on the proposed approach. In section 3.2, we explore to link the attackers' intentions and their strategies with STPA control loops of the system. Section 3.3 relates the ADTrees analysis and the attack surface of the system. In Section 3.4, we suggest extending this approach to using attack profile to focus the analysis of attack strategies.

3.1. Integrating Attack Modeling to the STPA Process

Attack modeling for system security analysis is an approach on top of STPA. Figure 1 illustrates the attack modeling process. The approach extends STPA process with three main steps which are the identification of attack profiles, the identification of unsecure control actions and the refinement into attack strategies. The attack profiles are defined to focus the analysis on specific attacker's capabilities.

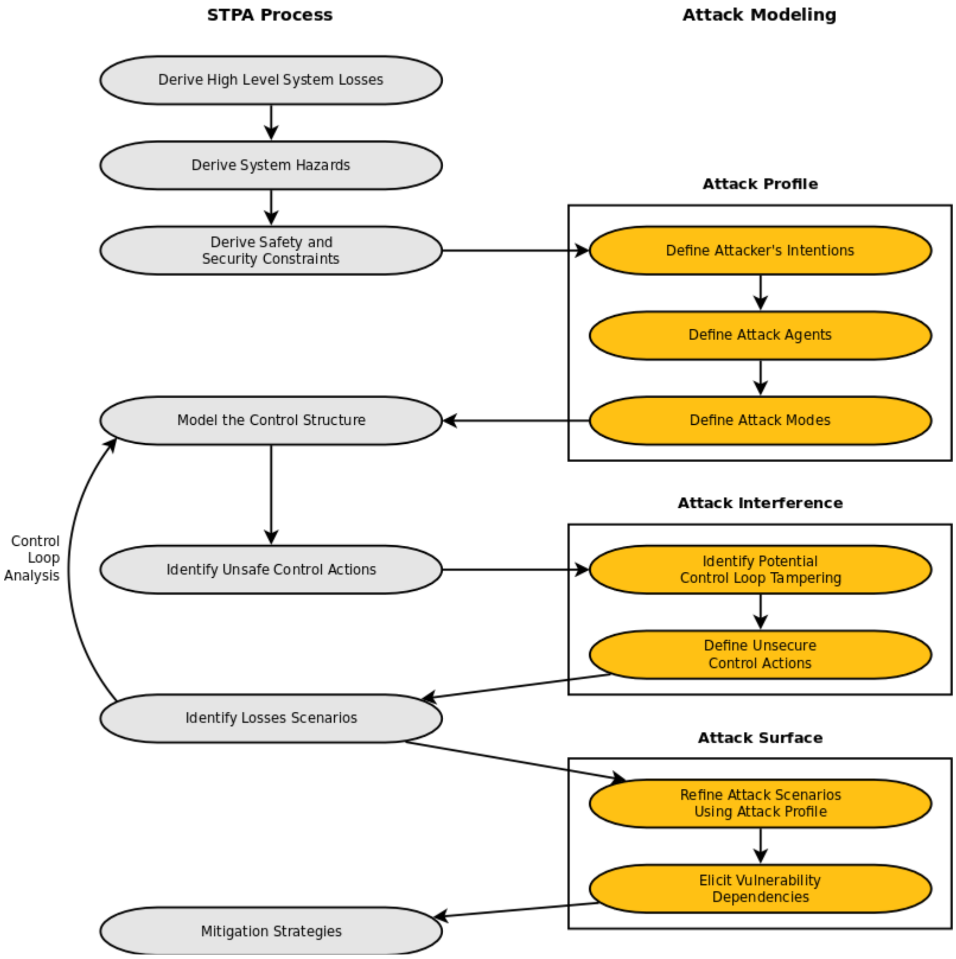

STPA Process
Attack Modeling
Derive High Level System Losses
Derive System Hazards
Derive Safety and Security Constraints
Attack Profile
Define Attacker's Intentions
Define Attack Agents
Define Attack Modes
Model the Control Structure
Control Loop Analysis
Attack Interference
Identify Unsafe Control Actions
Identify Potential Control Loop Tampering
Define Unsecure Control Actions
Identify Losses Scenarios
Attack Surface
Refine Attack Scenarios Using Attack Profile
Elicit Vulnerability Dependencies
Mitigation Strategies

Figure 1: Extension of STPA Process (Leveson & Thomas, 2013) with Attack Modeling

3.2. Bridging STPA Control Loops and Attack Defence Trees

In this section we explore how intention becomes action and how intention affects the STPA control loop. The difference between safety analysis and security analysis lies in the fact that the first does not consider the intention of an attacker. The attacker's intention and potential attack scenarios to accomplish the attacker's goal could be modelled using ADTrees with the root and upper attack nodes of the tree representing the attacker's intention and these nodes being decomposed further down the tree into specific attack steps.

To combine STPA and ADTrees modeling for security analysis, we suggest structuring the attack modelling following this pattern. Taking one attacker's aim as root of an ADTrees, we decompose it into more strategic intentions the attacker could envision to pursue the goal. We name this top part of the tree the *Intention Tier*. This part is composed only of attack nodes and is free of elements from the system's modeling.

From each resulting individual intention, we continue the ADTrees modelling by building subtrees which are now in the *Control Tier*. This step is done by considering the attacker's intention faced with the STPA control loop of the system. The attack node is systematically decomposed into sub attack nodes targeting or tempering with each element of the STPA control loop. We name this refinement between single intention attack node and its control-loop specific attack sub-nodes a *Tampering Chords* as its aim is to identify how an intention could tamper with one element of the control loop and eventually resonate with the entirety of the control loop. This represents how the attacker could tamper with the system's control and therefore trigger unsecured actions. Tampering Chords are the sets of connections between the Intention Tier and the Control Tier.

Different strategies or phases of an attack are analysed with regards to the STPA control loops of the system in the Control Tier of the ADTree. In the Control Tier the description of attack scenarios remains high level. The attack and defence nodes at the Control Tier are associated with an STPA control loop. Tampering Chords are the connections between the nodes in the Intention Tier and the Control Tier, they are the bridges to translating the attacker's intentions into specific disruptions to the system's STPA control loops. Tampering Chords are the key to establishing *unsecure control actions* triggered by the attacker's actions that need to be prevented.

The generic STPA control loop consists of four main elements which are Controller, Actuator, Controlled process and Sensor. The control model presented in STPA is meant to define unsafe control action for the controller and the control process. An attacker could tamper with any element of the control loop in a way that would trigger an unsecure control action by the system. Missing or inadequate actions in a control loop could be hazardous for the system.

Suspected system behaviour could be expressed as an intentional system failure (triggered by an attacker's action) and non-intentional system failure. The STPA causal factors can provide the rationale for how non-intentional system failures can occur. These could be complemented by ADTrees to give the rational for how intentional system failures can occur. We can use security constrains as countermeasures for the attack scenarios. Defence nodes are a shortcut for establishing security requirements.

3.3 Attack Surface and ADTrees

Individual attack nodes within the Control Tier are related to an element of the STPA Control Loop. These individual attack nodes might have sub defence nodes which will correspond to security constraints. They could also be refined further into more concrete actions. This refinement should reach a point where the attacker is exploiting a vulnerability of a component of the system to start or continue its attack. Such concrete attack actions are leaves in the ADTree. These leaves represent the attack surface of the system. The way they are combined gives the dependency between components' vulnerabilities. We name this part of the ADTree the *Component Tier* as it is closely related to the physical implementation of the system. Steps of concrete attacks in the Component Tier are combined to reach nodes of the Control Tier. By using these separate tiers, we distinguish the system surface of attack's exploit and the deception of the intelligence of the system by attacking its control loops.

Figure 2 illustrates how Tempering Chords seat at the boundary between Intention Tier and Control Tier, and how Components Tier refines the attack strategies by highlighting the attack surface vulnerabilities they exploit.

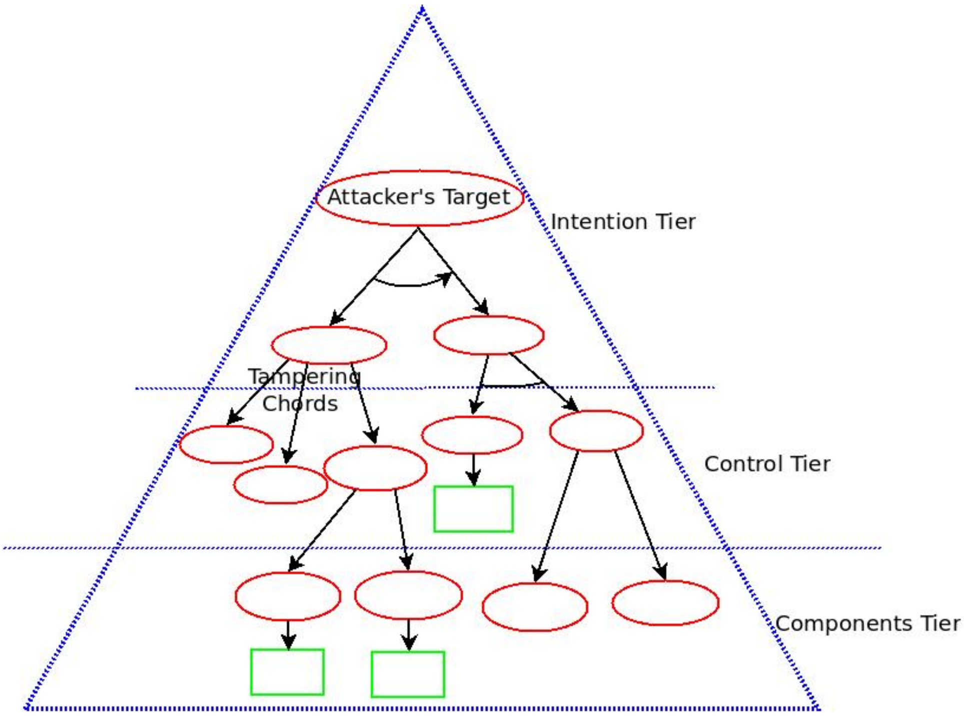


Figure 2: Tampering Chords and Ties of the ADTree Analysis

3.4. Enhancing attack scenarios by attaching Attack Profile

An *Attack Profile* is a way of expressing the attacker's abilities. Inspired by attack characterisations in (Schmittner, Ma, & Smith, 2014), we define attack profile as being is composed of two mains categories: the attack agent and the attack mode.

Attack agent is the abilities, knowledge and capabilities to attempt attack regardless of the intention. They could be categorised into script kiddie (attacker with a medium expertise, and he can apply self-learning material), blue hat (experience attacker who makes his attack with the purpose of showing his skills), black hat (experienced attacker who makes his attack for the purpose of terrorising, for money, for political ideals or religion motives) and elite hacker (expert designing and deploying their own tool to sell vulnerabilities that they discover in the black market). The capabilities of cyber-attacks are very high because the attackers have access to wide recourses and because such skills are related to intelligence.

Attack mode could be malicious, denial of service, spoofing identities and publish tools. Malicious code can be defined as a piece of code that usually connects to another program and can cause the system to behave unpredictably. Each code is designed for different reasons. The activation time depends on the design, for example trojans, worms and viruses. Their propagation is variable. User interaction may not be required like with viruses. Denial of Service: it is intensive connection from a group which aims to block the service provider and cause network congestion which lead to service delays. Spoofing Identities: is defined as a process in which a single computer, email, or other account associated with the service or a computer receive is hijacked or stolen by hackers. It necessitates some technique like fishing or social engineering.

Defining appropriate attack profiles and attaching attack profiles to ADTrees could help to focus the analysis by employing only capabilities related to the profile to narrow down the assessment. However, this approach should not prevent exploring wider attack profiles but helps to organise the analyses.

4. CASE STUDY

4.1 Cyber Attack Steel mill in Germany

The second known cyber-attack that resulted in a damage to physical systems concerned a German steel factory in December 2014. The Federal Office for Information Security announced the steel mill accident in their annual report without mentioning the name of the factory. Reportedly, the attackers used phishing email to gain access to the plant's network and then gain access to the production mill's network. The malware, which redirected to a malicious website, was downloaded to the targeted computer from a trusted email. The attacker was able to cause system components failure. This had a specific impact on the shutdown of critical components, which led to the impossibility of stopping the blast furnace (Lee, Assante, & Conway, 2014).

The steel mill was targeted with the intention to cause physical damage. The general network of the facility was hacked at the beginning of the attack. Then, the plant's production network which contains the management software of the steel mill was penetrated. The attacker took control of the plant's controlling system and succeeded in disabling the furnace's safety settings which caused serious damage to the infrastructure. According to the report, the attacker had a good knowledge and experience of the system. Figure 3 shows the design of the blast furnace's controlling system (Lee et al., 2014). The controlling system has a dashboard with several indicators such as the temperature of the furnace, its pressure, the water level in the tank. An operator has access to the dashboard and can require in an emergency the pumping of more water from the backup tank into the main tank or to stop hot blast and water bump. The computer of the controlling system controls the temperature of the furnace automatically by opening the blast furnace hot air valve and closing it. The cooling system is also controlled by the computer automatically using water pumps. The temperature in the furnace must be between 1500°C and 2000°C in order to produce steel.

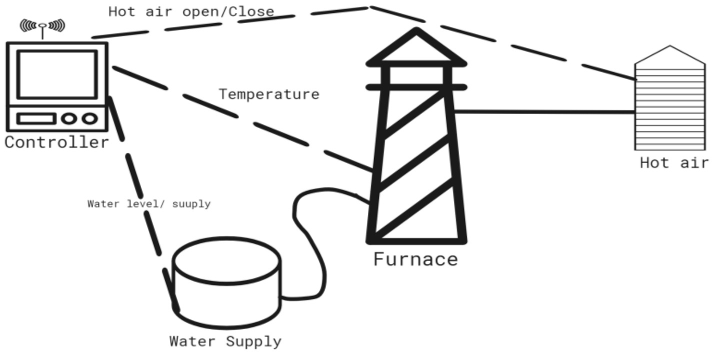


Figure 3: Steel Mill Simple Design System

4.2 STPA Analysis for Steel Mill Case Study

The first step in STPA is to define the purpose of the analysis, the system boundary, and losses and hazards for the system (see below).

Objective: to produce and sell steel
Losses:
L1- People die or injured in the steel mill.
L2- Steel mill production is stopped.
Hazards:
H1- Furnace is overheated [L1, L2]
H2- Furnace is unable to produce steel [L2]
H3- Furnace is physically injuring people [L1]
Safety constraints:
SC-1 Furnace temperature must be operated within limits [H1,H2,H3]
  SC-1.1 Furnace temperature must not exceed 2000C [H1,H2,H3]
  SC-1.2 Furnace temperature must not get lower than 1500C [H2]

The second step is to model the control structure. The analysis must identify the physical process and controllers, then define an unsafe control structure. Figure 4 shows the model of the control structure for the cooling mechanism, the heating of the furnace and their interactions.

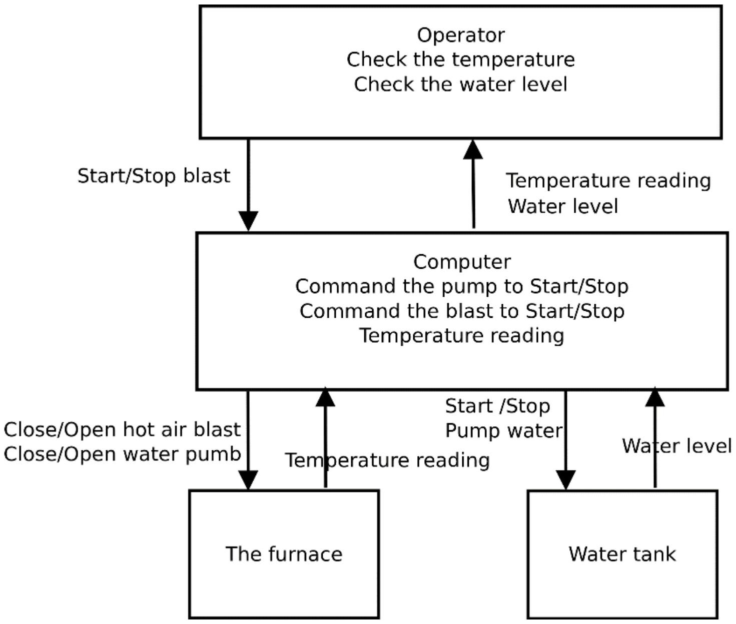


Figure 4: Steel Mill STPA Simple Control Loop

The third step in STPA is to identify unsafe control actions from the control structure which is mainly to find the behaviour to be prevented. Table 1 gives the system's unsafe control actions.

Table 1: STPA Unsafe Control Actions

	Not providing causes hazard	**Providing causes hazard**	**Too early, too late, order**	**Stopped too soon/Applied too long**
Open water pump	UCA-1: Computer does not provide open water valve when hot air valve close		UCA-2: Computer provides open water pump more than X seconds after hot air open	UCA-3: Computer stops providing open water pump too soon before the hot air valve fully open
Close water pump		UCA-4: Computer provides close	UCA-5: Computer provides close	

		water pump while hot air open [H1,2]	water pump more than X seconds before hot air close	
Open hot air		UCA-6: Computer provides open hot air while water pump is closed [H1,2]	UCA-7: Computer provides open hot air more than X seconds before water pumps open	
Close hot air	UCA-8: Computer does not provide close hot air when water pump is closed		UCA-9: Computer provides close hot air more than X seconds after water pumps close	UCA-10: Computer stops providing close hot air too soon before water pump is closed

Therefore, we can establish safety constraints (see below) from these unsafe control actions.

SC-1: The computer must not supply the open water valve when the hot air valve closes [UCA-1]
SC-2: The computer must not supply the open water pump for more than X seconds after opening the hot air [UCA-2]
SC-3: The computer must not supply the open water pump too early before fully opening [UCA-3]
SC-4: The computer must not supply a closed water pump when hot air is open [UCA-4]
SC-5: The computer must not supply the water pump closed more than X seconds before the hot air closes [UCA-3]
SC-6: The computer must not supply open hot air while the water pump is closed [UCA-6]
SC-7: The computer must not supply open hot air for more than X seconds before the water pumps open [UCA-6]
SC-8: The computer must supply hot air nearby when the water pump is closed [UCA-8]
SC-9: The computer must not supply hot air closed more than X seconds after the water pumps are closed [UCA-9]
SC-10: The computer must not interrupt the supply of hot air nearby too soon before closing [UCA-10]

The last step in STPA is to identify loss scenarios. This step is to explain how unsafe system behaviours could occur. For these scenarios, we consider multiple potential unsafe control actions. The updated model of Figure 5 includes the unsafe control action with the generic control diagram in blue. In Figure 6, the process model in red indicates what the controller believes. The process model for the water level indicates that the controller is to pump water from the reserve tank or use the backup pump when the water level is low. The temperature is normal. Thus, we need to redefine the process model in such a way that the computer should generate an alarm whenever the water level is getting low, helping the operator to send the command to stop the hot air or choose to do it manually.

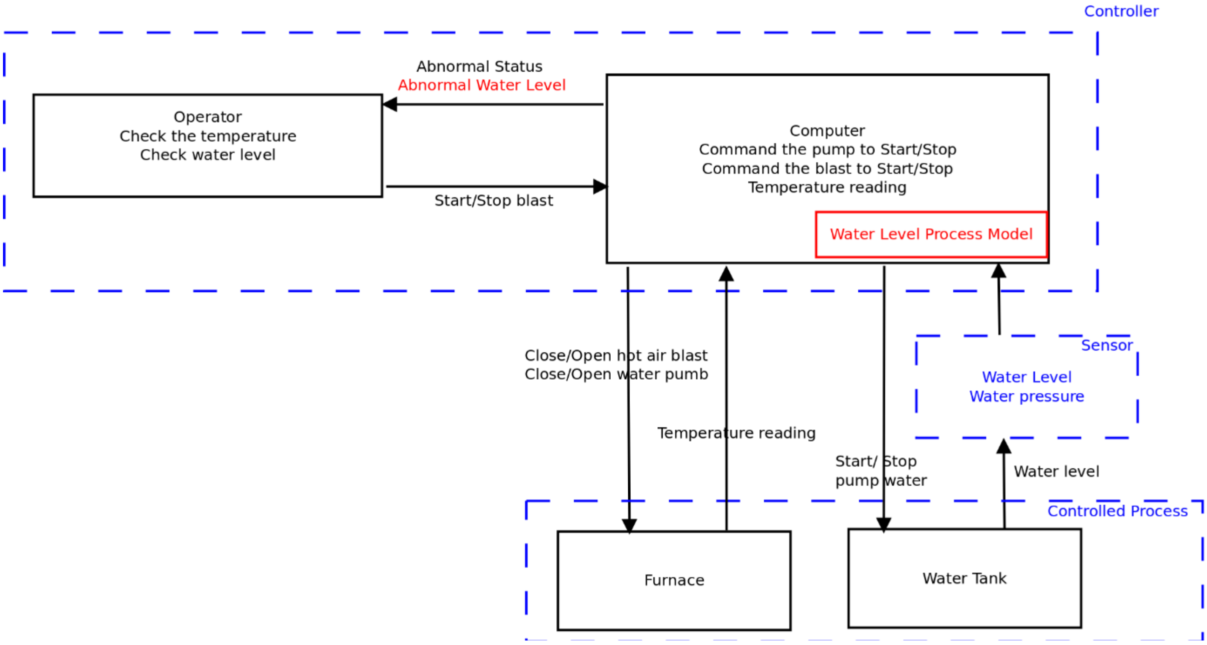


Figure 5: STPA Control Loop with Process Model

S-1: The operator did not recognize the rapid increase in the temperature indicator because the indicator showed normal status.

S-2: The operator responses to the water level decreases by pumping more water into the cooling system and switching the backup pump.

S-3: The rapid increase of the temperature leads to water leak; which results in more water being pumped to the cooling system; which results in the mixing of water and iron; which leads to the explosion.

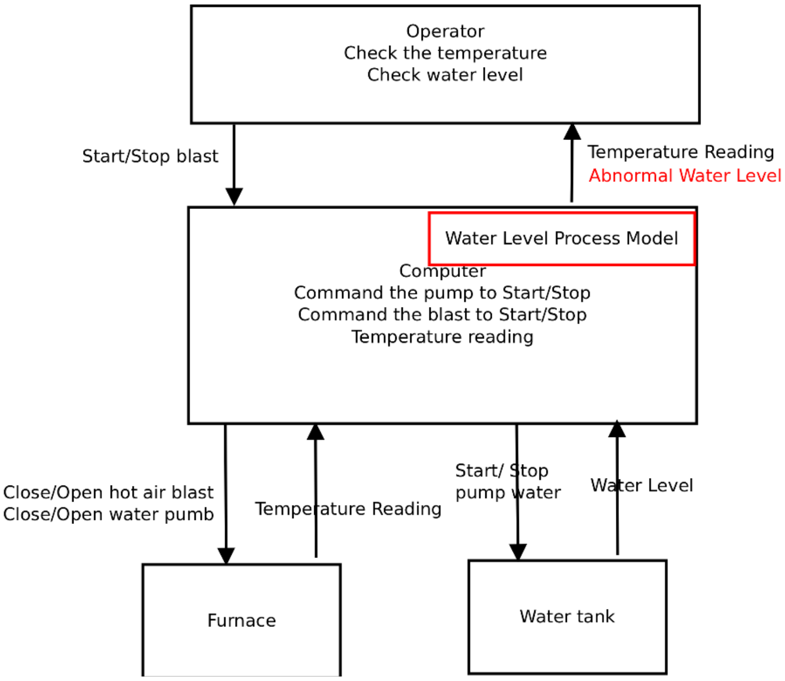


Figure 6: STPA Water Level Process Model

4.3 Steel Mill Attack Modeling

In this section we use ADTrees to model potential attacks in relation with the STPA analysis. In this case study, the intention of the attack is to cause life losses or enormous physical damage.

We build our ADTree, see Figure 7, following the steps of Section 3. The goal of the attacker is decomposed into intentions. Here, to simplify the tree, we showed a single intention. We then consider how this intention can tamper with the control loop of the system. We created three attack nodes as children of the intention attack node. These three nodes, which are unsecure control actions (USECA), could be later refined into specific attack sequences. In the example of Figure 7, the leftmost sub-tree corresponds to the successful attack described in the report. They exploited vulnerabilities in the networks and operating system (the sub-tree reaches elements of the attack surface). The rightmost sub-tree shows an example of a defence node.

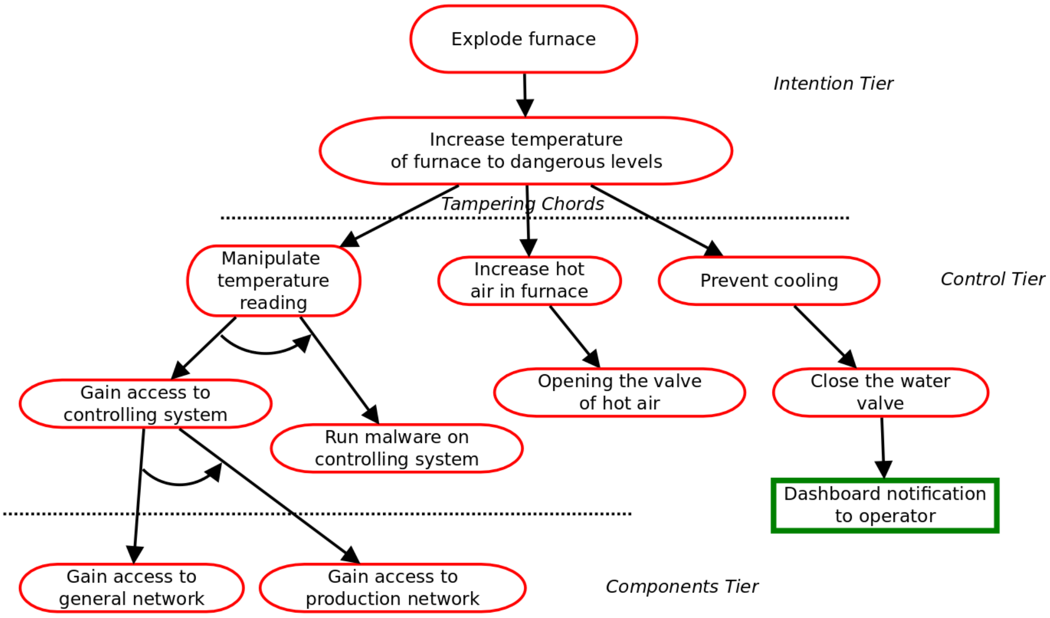

Explode furnace
Intention Tier
Increase temperature of furnace to dangerous levels
Tampering Chords
Manipulate temperature reading
Increase hot air in furnace
Prevent cooling
Control Tier
Gain access to controlling system
Run malware on controlling system
Opening the valve of hot air
Close the water valve
Dashboard notification to operator
Gain access to general network
Gain access to production network
Components Tier

Figure 7: ADTrees Scenario with vulnerabilities dependencies

The following USECAs correspond to the attack scenarios of Figure 7.
USECA-1 Attacker manipulates temperature reading
USECA-2 Attacker increases hot air in furnace
USECA-3 Attacker prevents cooling

For each USECA we could derive scenarios of attack from the ADTree, for instance the following scenario.
USECA-1
Components: Sensor
Control action 1: Gain access to controlling system
Control action 2: Gain access to general network
Control action 3: Gain access to production network
Control action 4: Run malware on controlling system

The attack leaves of the tree correspond to the vulnerabilities of the system. Note that an attack scenario would exploit individual vulnerabilities. The scenarios of attack indicate how these exploits could be combined. Such scenarios are therefore building a set of vulnerability dependencies which map the attack surface of the system.

The scenarios modelled with ADTrees can be refined using attack profiles. For instance, exploiting the networks and operating system vulnerabilities could well be done by attackers with different levels of expertise (e.g. script kiddie or elite hacker) which will make steps in the attacks to be more or less likely to take place. Their intentions might also differ. Attaching these attack profiles to the ADTrees could focus the analysis by bringing additional realistic aspects to the scenarios of attack.

5. CONCLUSION

In conclusion, we explored the modeling of attack strategies together with control structure of STPA using ADTrees. This should facilitate the elicitation of vulnerabilities most likely to cause harm to the system, and to define attack countermeasures. We propose to guide the building of ADTrees by scrutinising the way attacker's intention meet the control loops of the system. We also suggest using attack profiles to produce capability-focused attack scenarios. We applied this approach on a case study. We believe that this example shows the potential to help narrow down the attack scenarios modelled with the help of attack profiles. The connection between the STPA control loop and ADTrees elements offers a perspective in the design of modeling tools to establish unsafe actions in STPA, including the attacker's intention. This work is still in progress. We are developing prototype modeling tools to evaluate the implementation and effectiveness to help assess in the security of complex systems.

REFERENCES

Friedberg, I., McLaughlin, K., Smith, P., Laverty, D., & Sezer, S. (2016). STPA-SafeSec: Safety and security analysis for cyber-physical systems. *Journal of Information Security and Applications*. https://doi.org/10.1016/j.jisa.2016.05.008

Jhawar, R., Kordy, B., Mauw, S., Radomirović, S., & Trujillo-Rasua, R. (2015). Attack Trees with Sequential Conjunction. *ICT Systems Security and Privacy Protection*, 339–353. https://doi.org/10.1007/978-3-319-18467-8_23

Kordy, B., Mauw, S., Radomirović, S., & Schweitzer, P. (2014). Attack–defense trees. *Journal of Logic and Computation*, *24*(1), 55–87. https://doi.org/10.1093/logcom/exs029

Kriaa, S., Bouissou, M., & Laarouchi, Y. (2015, October 20). *A Model Based Approach For SCADA Safety and Security Joint Modelling: S-cube*. https://doi.org/10.1049/cp.2015.0293

Lee, R. M., Assante, M. J., & Conway, T. (2014). German steel mill cyber attack. *Industrial Control Systems*, *30*, 62.

Leveson, N., & Thomas, J. (2013, August). *An STPA Primer*. Retrieved from http://sunnyday.mit.edu/STPA-Primer-v0.pdf

Schmittner, C., Ma, Z., & Smith, P. (2014). FMVEA for Safety and Security Analysis of Intelligent and Cooperative Vehicles. In A. Bondavalli, A. Ceccarelli, & F. Ortmeier (Eds.), *Computer Safety, Reliability, and Security* (pp. 282–288). Springer International Publishing.

Schneier, B. (1999). Attack Trees. *Dr. Dobb's Journal*. Retrieved from http://www.schneier.com/paper-attacktrees-ddj-ft.html

Young, W., & Leveson, N. (2013). Systems Thinking for Safety and Security. *Proceedings of the 29th Annual Computer Security Applications Conference*, 1–8. https://doi.org/10.1145/2523649.2530277