# Safety Verification for Autonomous Ships

**Børge Rokseth[1,*], Odd Ivar Haugen[2] and Ingrid Bouwer Utne[1]**
[1] Department of Marine Technology, Norwegian University of Science and Technology, Norway
[2] DNV GL, Group Technology & Research, Trondheim, Norway

## ABSTRACT

Autonomous and unmanned ships are approaching reality. One of several unsolved challenges related to these systems is how to perform safety verification. Although this challenge represents a many-faceted problem, which must be addressed at several levels, it seems likely that simulator-based testing of high-level computer control systems will be an important technique. In the field of reliability verification and testing, design verification refers to the process of verifying that specified functions are satisfied over the life of a system. A basic requirement for any autonomous ship is that it has to be safe. In this paper, we propose to use the Systems-Theoretic Process Analysis (STPA) to (i) derive potential loss scenarios for autonomous ships and safety requirements to prevent them from occurring, and (ii) to develop a safety verification program, including test cases, intended to verify safety. Loss scenarios and associated safety requirements are derived using STPA. To derive a safety verification program, these unsafe scenarios and safety requirements are used to identify key variables, verification objectives, acceptance criteria and a set of suitable verification activities related to each scenario. The paper describes the proposed methodology and demonstrates it in a case study. Test cases for simulator-based testing and practical sea-trials are derived for autonomous ships. The case study shows that the proposed method is feasible as a way of generating a holistic safety verification program for autonomous ships.

**Keywords:** Autonomous Ships; Safety Verification; STPA; Maritime Safety; Test Case.

## 1. INTRODUCTION

Autonomous ships soon may become a reality, but there are many unsolved challenges, related for example to legal issues and safety (Levander, 2017). According to Rødseth & Nordahl (2017), one of the main premises for autonomous ships is that they must be safer than manned ships. An initial study on the new role of the human operator in autonomous ships shows that the potential for flawed decision making is still significant (Ramos, Utne, Vinnem, & Mosleh, 2018). Allocating functions to software controllers that previously were allocated to human operators may contribute to increased likelihood of human error because it may compromise the operator's situation awareness and system understanding (Hogenboom, Rokseth, Vinnem, & Utne, 2017).

Even if the computer control systems traditionally have performed tasks reliably, it does not necessarily mean that they will perform new and more complex tasks currently allocated to human operators in a reliable manner. New types of systems and increased complexity may introduce new failure modes and unforeseen system interdependencies (Utne, Sørensen, & Schjølberg, 2017). Wróbel, Montewka, & Kujala (2017) evaluated the impact of unmanned vessels on maritime transportation safety by analyzing previous conventional ship accidents and assessing what could have been different if the involved ship had been autonomous. The results indicated that many of the accidents might have been avoided with unmanned ships. On the other hand, crewmembers may have prevented innumerable accidents in innovative ways that an unmanned ship would not have been able to prevent.

* Corresponding author: +47 922 42 159, borge.rokseth@ntnu.no

An autonomous ship system may only be able to respond with predetermined decision logic whereas a bridge crew can improvise. Hence, it becomes even more important to foresee any potential hazardous situation, which may be encountered by the autonomous ship, and to embed appropriate responses into the system design. This makes system verification challenging, and it becomes necessary to design for verification. It may be, for example, advantageous to be able to determine specifications for simulation model prototypes at an early design phase and to determine how such models can be used to test assumptions and to explore potential emerging system behaviour.

In this paper, we propose and demonstrate a method using the Systems-Theoretic Process Analysis (STPA) to derive a safety verification program. The method is designed for application in the early design phases of a system. A safety verification program refers to a set of verification activities to be conducted to demonstrate that various potentially hazardous scenarios, which may violate safety constraints, will not take place. Verification activities range from formal model-based testing, simulator-based tests, practical trials, inspections, code review, documentation review, and more. The proposed method specifies verification objectives and corresponding acceptance criteria along with suitable verification activities and specifications for each verification activity. The integration of the management of verification activities at an early stage in the design process is important to facilitate efficiency in the testing, analysis, and inspection of the system.

Other references have previously presented methods for integrating STPA into the system design and development process. Leveson (2011) discussed safety-guided / safety-driven design based on STPA. Thomas et al. (2015) presented a method, which can be used to perform hazard analysis in parallel with system design based on STPA. Fleming & Leveson (2016) presented a method for analysing future concepts to identify hazardous scenarios at an early stage and allowing safety-driven development of these early concepts.

Thieme, Utne, & Haugen (2018) used a systems engineering approach to assess whether current risk models are suitable for application or adaptation to autonomous ships, and found that STPA is a suitable method. Wróbel, Montewka, & Kujala (2018a) identified a control structure for remotely controlled ships and applied STPA based on the control structure. In a similar study, Wróbel, Montewka, & Kujala (2018b) analysed autonomous ships where a system control structure was identified, and STPA focused on two control actions. In Abrecht & Leveson (2016), STPA was applied to analyse a platform supply vessel engaged in a target escort operation, with emphasis on the operational aspects. Rokseth, Utne, & Vinnem, (2017) applied STPA to evaluate whether it is beneficial as a substitute or complement to the current Failure Mode and Effect Analysis (FMEA) approach in the dynamic positioning (DP) industry. Results showed that many relevant accident scenarios did not necessarily involve component failures and that the current reliability and redundancy focused perspective is too narrow. STPA has a broader system perspective, which captures more accident scenarios. Aps, Fetissov, Goerlandt, Kujala, & Piel (2017) addressed challenges related to maritime traffic management safety in the Baltic Sea by developing a hierarchical control structure model for ship navigation and traffic maritime traffic management and using this model in an STPA. Rokseth, Utne, & Vinnem (2018) proposed an approach to integrate STPA into the current risk analysis and verification process for DP systems by using STPA to formulate verification objectives. The present paper is a further extension of this method and focuses on how STPA can be used to specify complete verification programs including acceptance criteria and specifications for how to conduct each specified verification activity.

The proposed method is not the first method to employ STPA as a foundation for verification. Abdulkhaleq & Wagner (2016) presented a risk-based approach for deriving formal test cases, where STPA was combined with model-based testing. This work is well suited for conducting formal verification of safety-critical software while retaining a holistic system safety perspective. In this paper, however, we propose a method, based on a similar approach, which is suitable for developing a safety verification program, including test cases for systems and concepts at different stages during system development. The method proposed in this paper represents a complementary approach to the method presented in Abdulkhaleq & Wagner (2016).

The proposed method in this paper is demonstrated through a case study for a general autonomous ship concept. The case study demonstrates how the proposed method can be applied at an early design stage to refine the concept or design, while simultaneously deriving a verification program, including test cases for simulator-based testing and practical sea trials.

## 2. METHOD

### 2.1. STPA

STPA is a hazard identification model based on the Systems-Theoretic Accident Model and Processes (STAMP), according to which accidents are caused by inadequate enforcement of control. Thus, safety becomes a control problem, rather than a question of avoiding individual failures and errors. The hierarchical control structure of the system under consideration is modelled by identifying how each layer of control enforces control on the next layer. STPA is then used to identify potentially Unsafe Control Actions (UCAs) and scenarios in which they may occur.

In Leveson & Thomas (2018), STPA consists of four steps. In the first step, the purpose of the analysis is defined by specifying accidents and system-level hazards. In the second step, the system under consideration is modelled as a hierarchical control structure. This includes identifying the functions of the various controllers and the relationship between controllers by specifying control action and feedback. In the third step, UCAs are identified. This is achieved by examining how each control action can lead to hazardous states by (i) not being provided, (ii) being provided, (iii) being provided too early or too late or in the wrong order, or by (iv) being provided for too long or too short. In the fourth step of STPA, scenarios that may cause the unsafe control actions to take place, or which may result in improper (or no) execution of control actions, are identified by examining relevant parts of the control loop that reflects specific parts of the previously modelled hierarchical control structure.

### 2.2. The Proposed Method

The method presented below should be initiated at the earliest stages of a system design process to ensure that the appropriate verification activities, such as simulator-based tests, practical trials, inspections and analyses, are facilitated as the design is developed. This may potentially both reduce the costs associated with verification and increase the achievable level of system confidence that may be gained through verification. The method consists of four steps:

- **Step 1:** Conduct an STPA analysis and derive safety requirements. The analysis should be refined together with the system design. When the analysis identifies needs, refinement can be achieved by refining the STPA related to specific functions to satisfy those needs. It is also possible to refine the analysis by recursively identifying scenarios in which safety constraints can be violated and new safety constraints for those scenarios of violation (Leveson, 2011).
- **Step 2:** Select a loss scenario and associated safety constraint and use them to identify key variables, verification objectives (aims) and acceptance criteria for the verification objectives.
- **Step 3:** Based on the key variables and formulation of the verification objectives, determine suitable means (verification activities) for satisfying the verification objectives. These may, for example, consist of a combination of simulator-based tests, practical trials, analyses, reviews and inspections.
- **Step 4:** Describe the setup and execution and, if necessary or possible, more concrete acceptance criteria for each test, analysis, review or inspection.

The output from the first step is a set of loss scenarios and safety requirements. Step 2 provides the foundation for describing the setup, execution and acceptance criteria for each verification activity. In Step 3, suitable types of verification activities are selected. Different types of

verification objectives will require to be addressed by different types of verification activities, and some must be investigated using several methods to achieve sufficient confidence. Step 4 produces specifications for conducting verification activities, such as test cases.

## 3. CASE STUDY

In this case study, we consider a ship that is capable of autonomous navigation, following a pre-planned trajectory, which may be updated in real-time. The trajectory consists of a set of waypoints and a travel speed between each waypoint. The ship is supervised or controlled by a remote operator in a shore control centre (SCC), as described, for example, by Rødseth, Kvamstad, Porathe, & Burmeister (2013), while there also is an automation system aboard the ship capable of autonomous navigation. In addition to deriving a safety verification program, the case study demonstrates how the proposed method can be used to analyse a design at an early phase, and how it can be used to refine the design by recommending specific functionality to be implemented to achieve a safe system.

### 3.1. Step 1: Conducting the STPA

#### 3.1.1. Engineering Foundation

The autonomous ship system architecture described by Rødseth, Tjora, & Baltzersen (2014) includes the following operational system modes: autonomous execution, autonomous control, direct remote control, indirect remote control, and a fail-to-safe mode. Under autonomous execution, the automation system executes a predefined plan while in the autonomous control mode, the automation system can update the nominal trajectory and make other changes to the system. When under indirect remote control, SCC can control the ship by updating the nominal trajectory, and indirect remote control mode, the heading and speed of the ship are controlled directly by means of a joystick. The fail-to-safe mode is considered a fallback strategy.

As also pointed out by Porathe, Hoem, & Johnsen (2018), ships do not really have a generally safe state. In the concept presented in Laurinen (2016), a fallback strategy is provided as a set of instructions to be executed, if necessary. An example is to hand over direct control to the remote operator. In this case study, we assume that the ship can be either in autonomous control mode, indirect remote control, or direct remote control. Fallback instructions can be provided and updated by the SCC before and during the voyage.

Figure 1 presents a possible control hierarchy for the autonomous ship. The remote operators in the SCC may, if the system is in the mode "indirect remote control", update the future trajectory based on real-time data, for example, to avoid a collision. This responsibility shifts to the Automatic Sailing System (ASS) if the system is in the mode "autonomous control". The autopilot contains the nominal trajectory, which may be updated by the SCC or the ASS, and calculates speed and heading reference signals based on the trajectory and navigational feedback (i.e., position, heading, speed and turn-rate). A motion control system (MCS) takes in the speed and heading request and coordinates the efforts of individual motion actuators (e.g., rudders, thrusters and propellers) such that they together generate control forces which result in a ship motion that corresponds to the requested heading and speed. SCC can bypass the autopilot by providing speed and heading reference signals directly to the MCS, through a joystick, if the system is in the mode "direct remote control".

For the actuator system to be able to generate the required forces, sufficient amounts of power must be available from the ship’s power system. Both the SSC and the ASS may configure or reconfigure the ship power system. Depending on the type of power system, this may include starting and stopping generators, altering the electrical distribution, and changing strategy with respect to operating mode of energy storage devices, such as batteries.

The sensor module is responsible for providing the navigation states of the own ship, and to detect and track and classify potential obstacles (both stationary obstacles such as a lighthouse or

an oil platform, and moving obstacles such as a ship or drifting ice). The ASS uses this data to try to predict the future trajectory of potential obstacles.

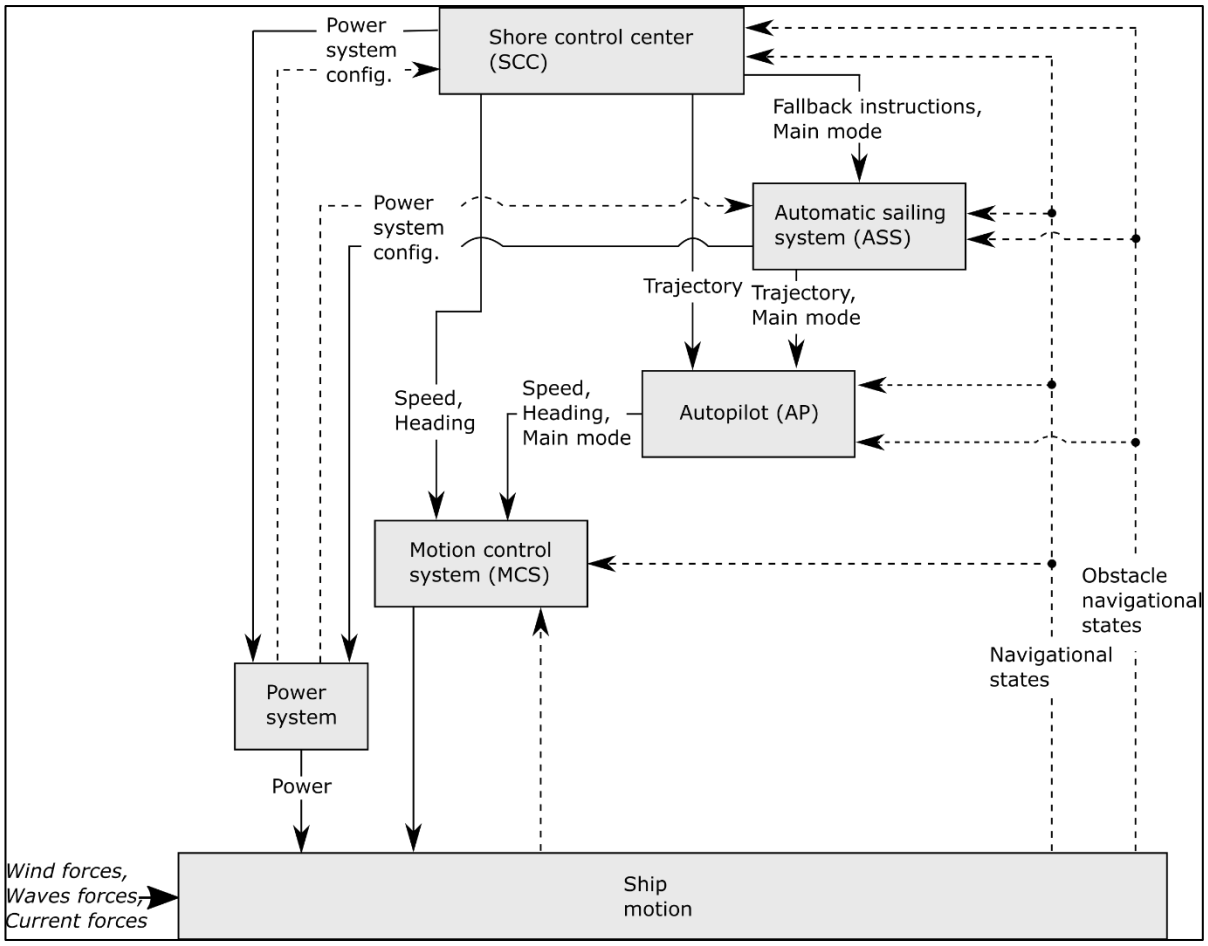


Figure 1: A possible control hierarchy for an autonomous ship

### 3.1.2. Accidents and System-level Hazards

In this case study, we have focused on accidents which may result in loss of human life, loss of mission, damage to property, and damage to the environment. We have chosen to focus on the following potential system accidents:

- A-1 The ship collides with a moving obstacle.
- A-2 The ship collides with a stationary obstacle, such as fixed structures or land.
- A-3 Loss of navigational control of the ship.

To identify and select system hazards which may lead to these accidents, we have reviewed the hazards presented in DNV-GL (2018), Wróbel, Montewka, & Kujala (2018a) and Wróbel, Montewka, & Kujala (2018b). We have then selected and reformulated the following three hazards for our case study:

- H-1 Ship violates minimum specified separation distance to a stationary or moving obstacle (A-1, A-2).
- H-2 Ship violation COLREG, or rule for sensible behaviour on the sea (A-1).
- H-3 Loss of (or reduced) ship manoeuvrability (A-1, A-2, A-3).

### 3.1.3. Unsafe Control Actions and Safety Constraints

In this case study, we present UCAs for a selection of control actions (see Table 1). We have focused on the two first modes of unsafe control (i.e., a control action that is required for safety is not provided, and an unsafe control action is provided). To formulate UCA-5 and UCA-6, we introduced the term "worst case single failure (WCSF)". This refers to the single failure, which, if it occurs under the current circumstances, will result in the worst consequences measured in terms of a specific type of consequence (e.g., loss of manoeuvrability).

Table 1 Examples of unsafe control actions

| Control action | Not provided | Provided |
|---|---|---|
| Trajectory (SSC to AP) | **UCA-1**: The navigational trajectory is not updated by SCC to avoid obstacles or ensure COLREG compliance when the ship is under indirect remote control. (H-1, H-2) | **UCA-2**: A navigational trajectory, which is such that loss of manoeuvrability will result in violation of minimum specified separation distance to stationary or moving obstacle, is provided. (H-1, H-2) |
| Fallback instructions (SSC to ASS) | **UCA-3**: Fallback instructions are not provided before the ship loses contact with the SSC while the ship is under indirect remote control. (H-1, H-2, H-3) | **UCA-4**: The fallback strategy "ask operator to take manual control" is initiated while the ship is executing a collision avoidance strategy. *Rationale: The remote operator, when handed direct control, is not able to find a trajectory that does not violate minimum specified separation distance to stationary or moving obstacle or COLREG.* (H-1, H-2) |
| Power system config (SCC to Power system) | **UCA-5**: The power system and actuator system is not reconfigured to a sufficiently robust state when i) the WCSF can result in loss of, or reduction in, manoeuvrability and ii) the ship is in a navigational situation where loss of, or reduction in, manoeuvrability may result in violation of minimum specified separation distance to stationary or moving obstacle. (H-1, H-3) | **UCA-6**: The power system or actuator system is reconfigured to a state where the WCSF can result in reduced or complete loss of manoeuvrability when the ship is in a scenario where loss of or reduced manoeuvrability may result in violation of minimum specified separation distance to stationary or moving obstacle. (H-1, H-3) |

Two example safety constraints for UCA-2 are:

- SC-2.1: The trajectory must be updated to avoid that a WCSF will result in violation of the minimum specified separation distance to stationary or moving obstacles, or alternatively.
- SC-2.2: The power system and actuator system must be reconfigured to a state where a WCSF will not result in the specified violation.

We proceed by looking closer at a scenario describing how the safety constraints defined above for UCA-2, may be violated. One such scenario of violation (SoV) is:

- SoV-2: The remote operator incorrectly believes that the manoeuvrability after a WCSF will remain sufficient to avoid violation of the minimum specified separation distance to stationary or moving obstacle in the event of the WCSF, and therefore provides the trajectory.

To make sure that this scenario does not take place, the following safety constraint should be enforced (SoV-SC refers to a safety constraint aimed at ensuring that a scenario of violation does not take place):

- SoV-SC-2.1: The autonomous ship must include a function to automatically assess whether the nominal trajectory will result in a violation of the minimum distance of separation in the event that the WCSF should occur, and to keep the remote operator informed of the consequences of a WCSF. As specified in the safety constraint under UCA-2; if a potential violation is detected, an alternative trajectory should be identified, or the power system and actuator system should be reconfigured to a state where a WCSF will not result in the detected violation.

We proceed by formalizing the idea of a system to perform such assessments automatically.

### 3.1.4. The Online Consequence Analysis System

The STPA analysis has pointed our attention towards a requirement for outfitting autonomous ships with a system that, according to our derived requirements, should automatically assess whether or not a provided trajectory will result in a violation of minimum distance of separation in the event that the WCSF should occur. Currently, similar functions to determine the consequences in terms of station-keeping capability are common for DP vessels (Rokseth, 2018). The operators on board these vessels are warned if station-keeping capabilities of the vessel will no longer be adequate after a WCSF, and they will then consider whether it is possible to reconfigure the system to a state where the WCSF will not result in inadequate station keeping capabilities. If this is not possible, they may have to abort the mission.

An online consequence analysis (OCA) system for providing similar functionality for autonomous ships has been proposed in Fossdal (2018). In this case study, we demonstrate how requirements and a verification program for such a system can be derived. We conceptualize the OCA system as consisting of the main parts, as illustrated in Figure 2. First, a set of WCSF candidates are defined. Exactly how this should be done depends on the specific ship under consideration. For a ship with a diesel-electric power system, loss of one entire electrical distribution including associated actuators may be a good candidate WCSF. Other examples may be the loss of control over the rudder or loss of a diesel engine. Then each candidate WCSF scenario is considered together with the current machinery configuration, power consumption and available power at each part of the power distribution to estimate what would be the remaining manoeuvring capacity. The result is a set of estimates on the “worst-case loss of manoeuvrability”-scenarios, which are finally simulated to assess the effect on the navigation capabilities of the ship at different points on the trajectory for a given prediction horizon. The consequences, in terms of a potential violation of the minimum specified separation distance to a stationary or moving obstacle, is evaluated for all potential scenarios to determine whether the trajectory is safe.

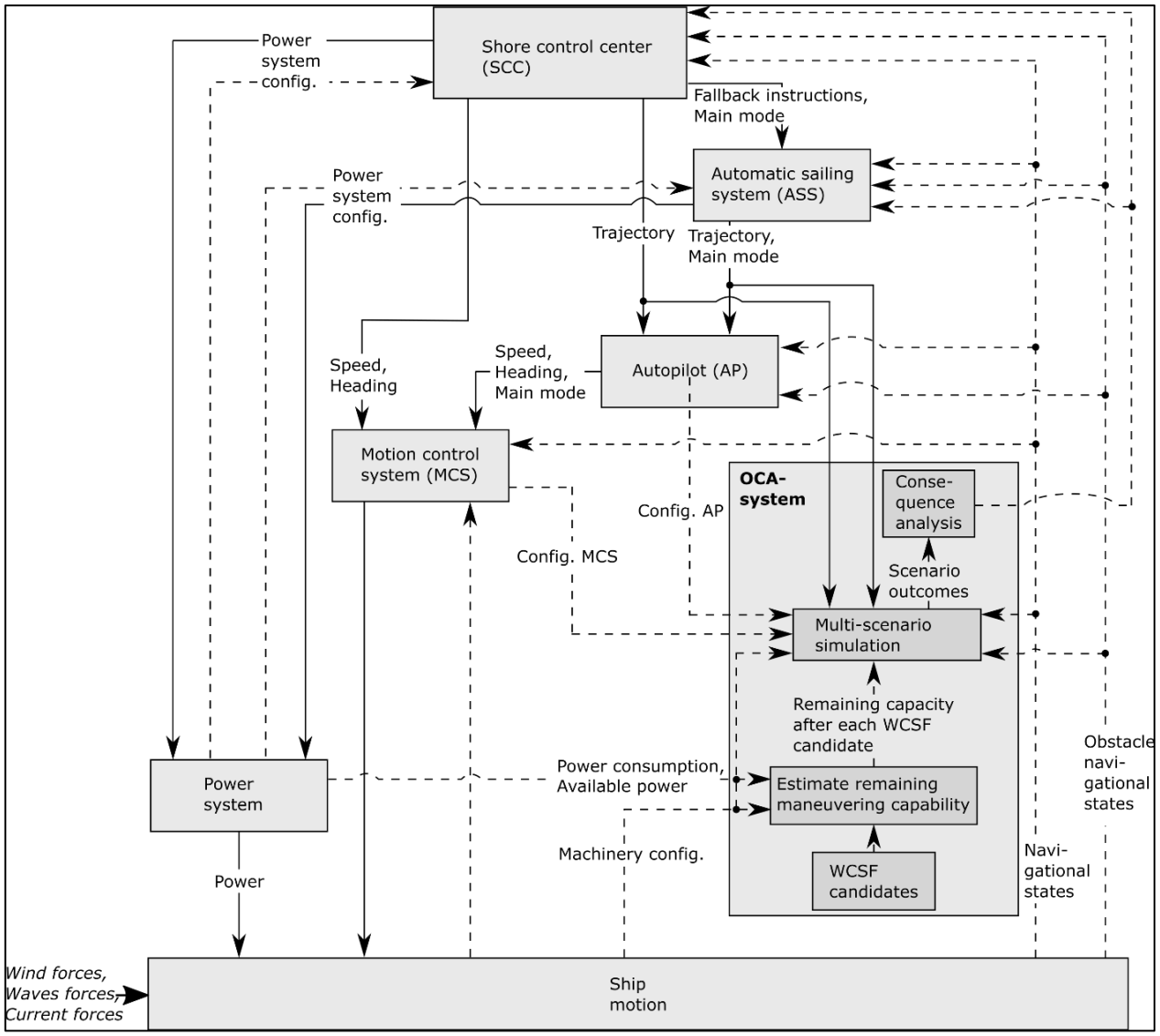


Figure 2: The OCA-system's role in the autonomous ship's control hierarchy

### 3.1.5. Refining the STPA for the OCA-system

In the previous section, we introduced the OCA to the autonomous ship control hierarchy (see Figure 2). Then we again performed an STPA using the same accidents and system-level losses as above, but this time also considering the OCA. An UCA associated with providing an affirmative or not on a given future trajectory based on whether or not it may cause a violation of the minimum distance of separation to an obstacle is:

- UCA-1: The OCA-system incorrectly provides an affirmative for a prediction horizon on a given future trajectory. Under the current system configuration and the prevailing weather conditions, the WCSF actually will result in violation of the minimum specified separation distance to stationary or moving obstacle.

Two safety constraints, which, if enforced, will ensure that this UCA does not take place, have been identified, together with a set of 28 potential loss scenarios related to the UCA. To limit the scope of this case study, we present the two safety constraints and one scenario of violation related to each:

- SC-1.1: Potential outcome of candidate WCSFs must be correctly estimated online for the given system configuration and set of circumstances.
    - SoV-1.1.1: The consequence analysis underestimates the outcome of a WCSF candidate in terms of loss of power or actuator capacity because the embedded logic is based on analyses of component failures that have disregarded relevant mechanisms, factors or circumstances that determine the outcome of the WCSF in question.
- SC-1.2: Correct predictions of the consequences of each potential loss scenario in terms of the ship's ability to avoid violation of the minimum distance of separation to obstacles, must be determined for each point on the given prediction horizon along the planned trajectory.
    - SoV-1.2.1: The multi-scenario simulation incorrectly predicts that violation of minimum distance of separation to an obstacle will be avoided for a given scenario because the SCC attempts to update the trajectory when the ship is in a control mode that does not support SCC trajectory update so that the autopilot rejects the update while the simulated autopilot does not.

These safety constraints and loss scenarios will serve as input to the next step in this case study.

### 3.2. Step 2: Deriving Key Variables, Acceptance Criterion and Verification Objectives

We proceed by identifying key variables, verification objectives and acceptance criteria for the two loss scenarios by inspecting the formulated safety constraints and scenarios. Key variables are variables that can be used to determine whether the loss scenario took place or not, and its severity. Typically, a loss scenario should be possible to express in terms of a set of specific combinations of values for the set of key variables. A general acceptance criterion could be expressed as a restriction on the relationship between such variables; see below.

#### 3.2.1. Scenario S-1.1

For this scenario, we find the following:

- Key variables
    1) The estimated power or actuator loss/limitation.
    2) The actual power or actuator loss/limitation.
- Verification objective:
    - Verify that all relevant mechanisms and factors that determine the power or actuator loss or limitations following each potential WCSF have been modelled in such a way that the power or actuator loss or limitations for any potential WCSF under any circumstance will not be underestimated.
- General acceptance criterion:
    - The estimated power or actuator loss should not be smaller than the actual power or actuator loss in any system configuration or circumstance.

#### 3.2.2. Scenario S-2.1

For this scenario, we find the following:

- Key variables:
    1) Reject/accept state of trajectory update in the real autopilot (cf. Figure 2)
    2) Reject/accept state of trajectory update in the simulated autopilot
    3) System operating mode

- Verification objective:
    - Verify that the reject/accept state of a trajectory update in the real autopilot is identical to the corresponding state in the simulated autopilot for a trajectory update from the SSC when the system is in the mode "automatic control".
- General acceptance criterion:
    - The reject/acceptance state in the two must be identical under any circumstance.

### 3.3. Step 3: Verification Activities

There are many types of verification activities available for collecting evidence regarding system conformity to requirements. Examples are simulator-based testing, practical tests (often referred to as practical sea-trials in the shipping industry), code review and documentation review. By inspecting the findings from Step 2, we can determine which types of activities are best suited to satisfy the verification objective.

#### 3.3.1. Scenario S-1.1

To satisfy this verification objective, it is necessary to test whether remaining manoeuvring capacity can be underestimated for any of the candidate WCSF. It may not be possible or safe to trigger all types of potential WCSFs physically. Therefore, we conduct tests on a high fidelity simulator instead. This also allows for rapid testing, and in turn more combinations of configurations and failures to be tested. It is also necessary to analyze the model used to estimate the consequences, or to review its documentation to determine i) whether all relevant mechanisms and factors have been considered and ii) to provide input to the simulator test to specify particularly interesting scenarios for example to test questionable assumptions. Finally, it is necessary to run a subset of the tests also in practical trials to gain confidence in the simulator model. Thus, review or analysis of documentation, simulator testing, and practical trials are necessary.

#### 3.3.2. Scenario S-2.1

To satisfy this verification objective, it is necessary to review the documentation of the autopilot and the simulated implementation to assess whether the same logic has been embedded in the two implementations. Findings should be confirmed in practical trials.

### 3.4. Step 4: Setup, Execution and Acceptance Criteria

In this step, the information from the previous steps is used to specify aim, setup, execution specifications and more specific acceptance criteria for each required verification activity.

#### 3.4.1. Scenario S-1.1

The three different verification activities identified for scenario S-1.1 is further specified in Table 2.

Table 2 Aim, setup, execution and acceptance criteria for scenario SoV-1.1.1

| | **Documentation review** | **Simulator test** | **Practical sea trials** |
|---|---|---|---|
| Aim: | 1) To verify that all relevant factors which may affect the performance of the system after the occurrence of a WCSF, have been considered and that those deemed relevant have been | -To compare the output of the model for estimating the remaining manoeuvrability after WCSFs with results from the high fidelity simulator model for all WCSFs in as many system | -To verify or increase confidence in the results from the simulator-based test. |

| | **Documentation review** | **Simulator test** | **Practical sea trials** |
|---|---|---|---|
| | implemented in the model for estimating remaining manoeuvring capability.<br>2) To provide input to simulator tests and practical trials by taking note of any assumptions used to develop the estimates of the remaining capacity and consider how these can be tested. | configurations and circumstances as possible | |
| Setup: | -A system expert will go through the documentation for the model for estimating the remaining manoeuvrability after WCSFs to verify that all relevant aspects have been considered and to take note of any questionable assumptions, and if possible to use these to define especially interesting system configurations for the simulator tests and practical trials. | -Exploratory tests should be conducted in a high fidelity simulator. The simulator should include a model of the actuator system, the power system and the motion control system. The simulator model will represent the real world. The actual model for estimating the remaining manoeuvrability after WCSFs will predict the effect in terms of capacity and available power for each actuator (or group of actuators sharing a power distribution) in each case. This prediction will be based on input from the high fidelity simulator. The simulator must be capable of simulating all candidate WCSFs defined for the system. The high fidelity simulator does not need to include high-level controllers such as a guidance system. | -These tests are to be conducted during practical sea-trials for the ship. As many as possible of the candidate WCSFs shall be simulated or triggered aboard the real vessel. Maximal output for each actuator shall be logged. As the force exerted by each actuator may not be directly measurable, indirect measures such as torque, RPM, power consumption can be used. The same scenario shall be simulated with the high fidelity simulator from case 2 and predicted using the embedded model for estimating the remaining manoeuvrability after WCSFs. |
| Execution: | | -Initiate a simulation with a selected actuator system configuration and power system configuration, as well as a realistic speed reference to the motion control system. For one of the candidate WCSFs, use the embedded model for estimating the remaining manoeuvrability after WCSFs to predict the effect on the power and actuator system loss/limitation. Then simulate the same candidate WCSF event and compare the results of the prediction and the | -The ship shall sail in a straight line with a fixed speed when a candidate WCSF is triggered. Maximal output for each actuator remaining after the failure is noted. The same candidate WCSF is triggered in the simulation model under identical circumstances, and the maximum subsequent output of each actuator is compared with the results from the practical test. |

| | **Documentation review** | **Simulator test** | **Practical sea trials** |
|---|---|---|---|
| | | simulation. This should be repeated for each candidate WCSF and for a set of combinations of system configurations. | |
| Specific acceptance criterion: | -Any mechanisms and factors, which may have a significant effect on power and actuator loss/limitation, must be accounted for realistically in the model for estimating the remaining manoeuvrability after WCSFs. | -The losses and limitations estimated by the embedded model for estimating the remaining manoeuvrability after WCSFs must be at least as severe as the simulated ones. | 1) The difference between the results from the practical trials and the simulation model should be reasonably small.<br>2) The automatic mechanisms and reconfiguration sequences for mitigating loss and restoring the system shall be identical in the simulator as those observed on the physical ship. |

### 3.4.2. Scenario S-2.1

The two verification activities found for this scenario is specified below in Table 3.

Table 3 Aim, setup, execution and specific acceptance criteria for scenario SoV-1.2.1

| | **Analysis and documentation review** | **Practical sea trials** |
|---|---|---|
| Aim: | To verify that the same logic is embedded in the simulated autopilot in the multi-scenario simulator in the OCA-system as in the real autopilot to accept or reject a trajectory update from the SCC based on the system mode | To verify that a trajectory update which is rejected by the autopilot is also rejected by the simulated autopilot in the multi-scenario simulator |
| Setup: | A system expert should review the documentation and/or code of the autopilot and the multi-scenario simulator model to verify that identical criteria are implemented to determine acceptance or rejection of trajectory updates with respect to operational mode in the two implementations of the autopilot. | The ship is sailing under autonomous control along some trajectory, with the OCA-system active. SCC will try to upload a new trajectory. Observe whether the trajectory is accepted or rejected by the autopilot and the simulated autopilot in the multi-scenario simulator. |
| Execution | | The test is to be conducted during practical sea trials |
| Specific acceptance criteria | The exact same criteria must be used in the two implementations. Alternative solutions which ensure that the loss scenario will not occur, such as reading the trajectory in the simulation model directly from the nominal trajectory in the autopilot, can be accepted. | After the test, the nominal trajectory shall be identical in the autopilot in the multi-scenario simulator and the real autopilot. |

## 4. DISCUSSION

There are at least two advantages of developing a safety verification program for autonomous ships at an early design stage. First, this enables us to design the system in such a way that it more readily can be tested and as such, may make it possible to gain a higher degree of confidence in the system. Further, verification may become less costly because design facilitates efficient and thorough testing, review, inspection etc. For example, being aware of the verification activity specified under "documentation review" related to the scenario SoV-1.1.1, a system developer may, in an orderly fashion, list the stated factors that have been included in the OCA-documentation to account for how a WCSF may affect the performance of the system, along with the rationale for including them. The developer can also list factors and mechanisms that have been considered (in terms of their potential for affecting the outcome of any WCSF candidate) but not included, as well as the rationale for not including them. Such input will not only help to point the developer's attention to matters that are important for safety and help the developer to approach them in a systematic manner. It might also reduce the cost of this particular verification activity (as opposed to the case where it is identified as necessary after the OCA-system has been developed). This is because the system documentation can be prepared and organized with verification in mind by making sure that the information required for verification actually can be found.

As another example, consider the need identified for verifying the high fidelity simulation model's ability to realistically simulate the effect of potential WCSFs (see “practical trials” under SoV-1.1.1). This requires means for artificially triggering potential WCSFs to be designed into the physical system (to facilitate comparison between the physical system and the high fidelity simulation model). Implementing such functionality is probably much easier to do if the need is defined at an early stage in the design process than it if were to be defined after the design was finished. Hence, it is important to identify the need at an early stage in the system design.

Furthermore, considering safety verification at an early stage makes it possible to develop virtual prototypes and digital twins also at an early stage because our analysis will define various needs and ways of utilizing such models. For example, based on this limited case study only, we know that the high fidelity simulation model to be used in the simulation-based tests for scenario SoV-1.1.1 should include realistic models of the actuator system, the power system and the motion control system, which are capable of realistically simulating each of the potential WCSFs.

Already at the early stage of the development of autonomous ships, OCA-systems (Fossdal 2018), as well as other concepts, such as anti-collision algorithms (Johansen, Perez, & Cristofaro, 2016), are being developed. There are, however, no means for developing requirements to simulation models for testing such emerging systems. The case study in this paper provides insight into this subject with respect to the OCA-system.

## 5. CONCLUSION AND FURTHER WORK

The method presented in this paper is intended to integrate system design development and the development and management of a safety verification program, including test-case generation. A case study has been conducted to demonstrate that the proposed approach is suitable for analyzing a design. The purpose is at an early phase to derive functionality that is required for safety and to also identify a safety verification program including test cases for the derived functions.

No matter how high the reliability is of the technical system of an autonomous ship, it is reasonable to assume that technical failures, which potentially may lead to a loss, will occur sooner or later. Even though efforts have to be put into prevention of such failures, e.g., in terms of thorough and efficient verification of the ship, it is necessary to design the ship in a resilient manner to accommodate failures and unplanned events without resulting in accidents. The case study conducted in this paper outlines one functionality (the online consequence analysis) that is intended to introduce such resilience to the autonomous ship design.

Many methods for hazard analysis will focus on one aspect or another related to safety, such as reliability, or one type of engineering disciplines, such as software or machinery. Our experience

is that STPA directs our attention to any type of issue or part of the system that requires our attention with respect to safety. This is important because verification activities should not be limited to only a certain kind of concern or engineering disciplines, such as reliability tests, or software testing. A verification program for system safety should include activities designed to verify a wide range of different aspects of a system, such as, for example, the safety of computer control systems, and their ability to handle different abnormal situations, and the system's ability to support human operators in obtaining the necessary situational awareness. While quite limited in scope, the case study presented in this paper demonstrates that a wide range of safety aspects and system parts can end up as the subject of safety verification as a consequence of the holistic safety perspective offered by STPA.

Based on the results from the case study, it is reasonable to expect that a more comprehensive analysis of the initial control structure model will reveal many additional safety constraints and functions that are required for safety, which have yet not been considered by the autonomous ship research community. Therefore, further work should focus on conducting a more comprehensive case study using the proposed method.

## ACKNOWLEDGEMENTS

This work is funded by the project Online risk management and risk control for autonomous ships (ORCAS). The Norwegian Research Council, DNVGL and Rolls Royce Marine are acknowledged as sponsors of project number 280655.

## REFERENCES

Abdulkhaleq, A., & Wagner, S. (2016). *A Systematic and Semi-Automatic Safety-Based Test Case Generation Approach Based on Systems-Theoretic Process Analysis* (Vol. V). Retrieved from http://arxiv.org/abs/1612.03103

Abrecht, B., & Leveson, N. G. (2016). *Systems theoretic process analysis (STPA) of an offshore supply vessel dynamic positioning system*. Boston, MA: Massachusetts Institute of Technology. Retrieved from http://hdl.handle.net/1721.1/104618

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*(Supplement C), 2–12. https://doi.org/https://doi.org/10.1016/j.proeng.2017.03.090

DNV-GL. (2018). *Class guideline DNVGL-CG-0264: Autonomous and remotely operated ships*.

Fleming, C. H., & Leveson, N. G. (2016). Early Concept Development and Safety Analysis of Future Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*, *17*(12), 3512–3523. https://doi.org/10.1109/TITS.2016.2561409

Fossdal, M. (2018). *Online Consequence Analysis of Situational Awareness for Autonomous Vehicles*. Norwegian University of Science adn Technology.

Hogenboom, S., Rokseth, B., Vinnem, J. E., & Utne, I. B. (2017). Human Reliability and the Impact of Control Function Allocation in the Design of Dynamic Positioning Systems. *Submitted to the Journal of Reliability Engineering and Safety Science*.

Johansen, T. A., Perez, T., & Cristofaro, A. (2016). Ship collision avoidance and COLREGS compliance using simulation-based control behavior selection with predictive hazard assessment. *IEEE Transactions on Intelligent Transportation Systems*, *17*(12), 3407–3422. https://doi.org/10.1109/TITS.2016.2551780

Laurinen, M. (2016). *Remote and Autonomous Ships: The next steps*. *AAWA: Advanced Autonomous Waterborne Applications*. Retrieved from http://www.rolls-royce.com/~/media/Files/R/Rolls-Royce/documents/customers/marine/ship-intel/aawa-whitepaper-210616.pdf

Levander, O. (2017). Autonomous ships on the high seas. *IEEE Spectrum*, *54*(2), 26–31. https://doi.org/10.1109/MSPEC.2017.7833502

Leveson, N. G. (2011). *Engineering a safer world: Systems thinking applied to safety*. Cambridge, MA: The MIT Press.

Leveson, N. G., & Thomas, J. P. (2018). *STPA Handbook*.

Porathe, T., Hoem, Å., & Johnsen, S. (2018). At least as safe as manned shipping? Autonomous shipping, safety and " human error ." In *Safety and Reliability–Safe Societies in a Changing World* (pp. 417–425). Trondheim.

Ramos, M. A., Utne, I. B., Vinnem, J. E., & Mosleh, A. (2018). Accounting for Human Failure in Autonomous Ship Operations. In *In Safety and Reliability–Safe Societies in a Changing World* (pp. 355–363). Trondheim, Norway.

Rødseth, Ø. J., Kvamstad, B., Porathe, T., & Burmeister, H. C. (2013). Communication architecture for an unmanned merchant ship. *OCEANS 2013 MTS/IEEE Bergen: The Challenges of the Northern Dimension*, (314286). https://doi.org/10.1109/OCEANS-Bergen.2013.6608075

Rødseth, Ø. J., & Nordahl, H. (2017). Definitions for Autonomous Merchant Ships, 22. Retrieved from http://nfas.autonomous-ship.org/resources/autonom-defs.pdf

Rødseth, Ø. J., Tjora, Å., & Baltzersen, P. (2014). *D4.5: Architecture specification*.

Rokseth, B. (2018). *Safety and Verification of Advanced Maritime Vessels: An Approach Based on Systems Theory*. Norwegian University of Science and Technology.

Rokseth, B., Utne, I. B., & Vinnem, J. E. (2017). A systems approach to risk analysis of maritime operations. *Journal of Risk and Reliability*, *231*(1), 53–68. https://doi.org/10.1177/1748006X16682606

Rokseth, B., Utne, I. B., & Vinnem, J. E. (2018). Deriving verification objectives and scenarios for maritime systems using the systems-theoretic process analysis. *Reliability Engineering and System Safety*, *169*(March 2017), 18–31. https://doi.org/10.1016/j.ress.2017.07.015

Thieme, C. A., Utne, I. B., & Haugen, S. (2018). Assessing ship risk model applicability to Marine Autonomous Surface Ships. *Ocean Engineering*, *165*(June), 140–154. https://doi.org/10.1016/J.OCEANENG.2018.07.040

Thomas, J., Sgueglia, J., Suo, D., Leveson, N., Vernacchia, M., & Sundaram, P. (2015). An Integrated Approach to Requirements Development and Hazard Analysis. *SAE Technical Paper*. https://doi.org/10.4271/2015-01-0274.Copyright

Utne, I. B., Sørensen, A. J., & Schjølberg, I. (2017). Risk management of autonomous marine systems and operations. In *ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering* (p. V03BT02A020-V03BT02A020).

Wróbel, K., Montewka, J., & Kujala, P. (2017). Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. *Reliability Engineering and System Safety*, *165*(August 2016), 155–169. https://doi.org/10.1016/j.ress.2017.03.029

Wróbel, K., Montewka, J., & Kujala, P. (2018a). System-theoretic approach to safety of remotely-controlled merchant vessel. *Ocean Engineering*, *152*(January), 334–345. https://doi.org/10.1016/j.oceaneng.2018.01.020

Wróbel, K., Montewka, J., & Kujala, P. (2018b). Towards the development of a system-theoretic model for safety assessment of autonomous merchant vessels. *Reliability Engineering & System Safety*, *178*(June), 209–224. https://doi.org/10.1016/j.ress.2018.05.019