Assessing the Trustworthiness of Manned and Unmanned Ships

T R Searle* MEng MIET

* *Frazer-Nash Consultancy, UK*
* Corresponding Author Email: t.searle@fnc.co.uk

Synopsis

One of the key challenges ship builders will face when introducing more complex control, and increasing autonomy, is managing the large volume of information that will be available to operators (who may be in a remote location) and providing a level of confidence in the correct operation of the system. Furthermore, ship operators will need to understand the implications of any unexpected or anomalous behaviour in terms of the impact on the capability and operability of the ship.

This paper presents a novel strategy for aggregating the large amounts of information, and modulating the information presented to the operator through an independent system of "trustworthiness" assessment.

The trustworthiness concept provides a means by which complex decisions can be more easily assessed and actioned in the face of multiple sources of evidence. Complex decisions are broken down into a hierarchy of factors, each of which considers its own sources of evidence and the implication of this evidence on distinct aspects of the ship. In some cases, the data constituting a given source of evidence might not be reliable – it could be noisy, partial, or completely absent. If left un-monitored, this could result in a control system making an incorrect decision based on missing or poor quality input data.

The ability to assimilate data whilst conveying any uncertainty or absence of data can guard against poor decision-making. In the presence of a multitude of sensors that all contribute towards a single decision, the trustworthiness concept can combine the outputs, consider the full breadth of the available information, and process them with limited human oversight requirement, to ultimately make more informed decisions in a more timely manner.

This calculated value of trust provides useful contextual information valuable for many different purposes. For example, it can be used to modulate the amount of intervention required by the operator, and the level of detail of information presented to them. It can also be used to adjust the size of an exclusion zone for an autonomous ship, to reduce the likelihood of collisions.

Our paper/presentation will describe the assessment process and the proposed structure of trustworthiness as applied to the marine industry, and shall provide the audience with examples of how this could be implemented in practice to safely reduce manning requirements on autonomous, or semi-autonomous ships.

Keywords: Trustworthiness; Autonomy; Self-assessment

Author Biography

Tom Searle Consultant at Frazer-Nash Consultancy in Bristol, UK. With a degree in Integrated Mechanical and Electrical Engineering, Tom is a Systems Engineer with a particular focus on the application of modelling and simulation, working primarily within the Defence and Naval sectors.

 http://doi.org/10.24868/issn.2515-8198.2019.002

1. Introduction

1.1. The Problem

This paper is a response to the Marine Electrical and Control Systems Safety Conference (MECCS) 2019 call, focusing on the safe interaction between people, processes, commercial standard equipment, and plant. Specifically, it tackles the topic of 'Trust' in lean-manned and autonomous ships, posing the question: 'can systems be trusted to do the right thing, at the right time?'

With the ever-increasing use of complex, interacting control systems aboard ships, ship builders and operators alike are faced with a challenge: how to best manage and utilise the wealth of information available from the network of systems and sensors distributed across the ship, and thus make more informed decisions in support of safe and effective operation.

On its own, the act of understanding sub-system operational data by operators and ship staff is non-trivial due to the number of different sources (including both quantitative and qualitative), and the volume of information they can produce. Sub-systems may use hundreds of measured parameters; monitoring these signals and identifying potentially problematic patterns or trends in each of them is non-trivial.

Furthermore, the act of understanding the implications of this sub-system level data on the ship's ability to operate as a whole system – and safely – is more difficult still. In an era of lean-manning and a move towards more complex and increasingly-automated (and even fully autonomous) ships, these issues will only be exacerbated further.

Whilst there are many potential applications for the concepts presented in this report, with reference to the above, the work presented here focuses on addressing two specific industry problems within the context of autonomy and trustworthiness:

- **Lean-Manning:** As the need for lean-manning becomes more prevalent, how can ships be operated effectively, efficiently, and safely given the conflicting requirement for more complex ships to be operated by fewer personnel?

- **Increased Shipping Density:** Globally, the volume of shipping traffic continues to rise, increasing the risk posed by and to ships at pinch-points such as narrow shipping lanes and ports. As ships are deployed into more densely populated operational areas, how can the safe operation and co-operation of these ships be ensured?

The work presented should not be considered a conclusive or complete solution to the above problems. Instead, it captures a proposition for a concept that, once matured, will help to address them. It is hoped that the outputs from this work will initiate further discussion and development to ultimately realise the Trustworthiness concept and its application to the maritime industry

2. Trustworthiness Concept

While there are many existing frameworks defining safety (and reliability, availability etc.) they have significant limitations when attempting to apply them to the concept of trustworthiness. This is because there is no one single attribute that defines trustworthiness: just because something is safe or reliable or available does not make it trustworthy. Instead, trustworthiness is the aggregation of many attributes such that a system can be expected to perform the requested action on demand.

A proposed definition of trustworthiness, and a way of quantifying it through the concept of 'trustworthiness levels', has been developed from existing definitions: this section outlines these definitions.

2.1. Trustworthiness Aspects

To offer a broad definition of what constitutes trustworthiness, and to provide a structured process for its assessment, it has been broken down into five 'aspects' inspired by the British Standards Institution Publicly Available Specification (PAS) 754 [1].

PAS 754 provides a definition of trustworthiness as applied to a software development and delivery context. There are similarities between the development of software and systems: they are both a complex integration of many simpler elements; they are both employed to deliver a complex effect 'greater than the sum of their parts'; and accordingly the design, development, test, and deployment are becoming increasingly difficult due to the need to understand complex emergent behaviours beyond the design boundary.

The PAS 754 trustworthiness aspects are therefore carried through and reused in this context, with some development and refinement of their definition to better align them to the domain context of marine systems and ships:

 http://doi.org/10.24868/issn.2515-8198.2019.002

Trustworthiness Aspect	Description
Safety	Ability to operate free from risk of causing unacceptable or intolerable levels of harm.
Reliability	Ability to operate as requested, without incurring a fault.
Availability	Ability to perform tasks when requested.
Resilience	Ability to function in case of fault, damage or error.
Security	Ability to function without inadvertent or malicious external influence.

Table 1: Trustworthiness Aspects Definition

These trustworthiness aspects provide the basis of a framework for assessing the effects of events (for example operational issues, external factors, or performance parameters) on the ship by assessing the way in which these events impact each factor, and therefore the ship's ability to operate.

The aspects are closely linked and are often interrelated, but are distinct enough to be considered separately. Importantly they provide value by encouraging consideration of system performance from different perspectives, and ultimately the aggregation of all of these factors into a single parameter in support of more informed but streamlined decision making.

A ship which meets all aspects of this framework can be described as:

"having the ability to perform a function, when demanded, without undesired influence or risk of causing harm".

In the context of this work, this would make the ship 'trustworthy'.

In practice, the relative importance of each aspect is likely to depend on the particular ship, system, environment, or mission - for example security may be considered relatively more important compared to other aspects depending on the cargo, or reliability may be considered relatively less important in open water or fair weather than in dense shipping traffic or stormy weather. An investigation to understand the potential philosophies, sensitivities, and usefulness of such weightings will form part of the future development of this concept of trustworthiness levels.

The degree of trust in a ship or system is split into five levels, denoting the level of confidence in its ability to perform as expected. These 'trustworthiness levels', similar in concept to the well-established Safety Integrity Level (SIL) or Development Assurance Levels (DAL), range from TL 0 (no trust) to TL 4 (complete trust) and provide an indication of the level of confidence an operator (or overseer, or remote controller etc.) can have in the ship (or system) functioning as expected when demanded.:

Trustworthiness Level	Description	Implication
TL 4	Complete trust in the system	The system may be expected to behave as specified when requested. Autonomous decision making can be expected to make sensible decisions based on all available data.
TL 3	High trust in the system	The controller may expect the system to behave as specified, but should be aware of sub-components or certain aspects operating with reduced trust, and should consider the suitability of using the system to perform tasks.
TL 2	Medium trust in the system	The system can be expected to perform a limited set of activities with increased operator oversight.
TL 1	Low trust in the system	There is a high risk that the system will be unable to perform in a suitable manner for the current scenario.
TL 0	No trust in the system	The behaviour of a system at TL 0 cannot be guaranteed in any way. Removal of all autonomous control should be attempted but this may not result in successful transfer of control from the system.

Table 2: Trustworthiness Level Definition - The system may be a subsystem within a ship, a whole ship, or a collection of ships operating in collaboration to perform mission objectives

 http://doi.org/10.24868/issn.2515-8198.2019.002

Within the maritime domain, the concept of trustworthiness is applicable to both the operation of a whole ship (manned, remote-operated, or autonomous) and the operation of individual system on-board a vessel. In the context of a 'maritime system', examples of the implications of these levels of trustworthiness are as follows.

A high TL conveys to the ship's captain, remote operator, or local equipment operator that the ship (or its subsystems) has no identified issues, its system health checks are positive, and there are no known impediments to it performing any task of which it is capable. A captain or the ship's staff would be confident in tasking the ship, while a remote overseer would be content with it operating autonomously in crowded, or safety-critical waters.

A medium TL could indicate that an isolated or non-essential sub-system's performance is suboptimal, or that ship function is affected by external effects such as inclement weather. An operator may continue to use the ship, but increase (manually or autonomously) safety margins to account for the drop in Trustworthiness, or remove it from theatre to reduce risk to the ship or to third-parties.

A low TL may indicate that an issue has been identified that directly affects the ship's ability to safely perform the tasks required of it. A low TL may also be caused by insufficient verifiable information being available to determine trust (attributable to either problems within the ship or losses during transmission). In this situation an operator might decide to continue with reduced or removed autonomy until it can be relieved, or they might pro-actively take mitigating actions to reduce the risk to the vessel and others. A remote overseer may choose to coordinate with other nearby ships to continue the task, or to provide a clear route to extract the ship safely.

3. Trustworthiness Assessment

While ultimately presented as a measure of whole-ship performance, it is important that the trustworthiness of the ship is informed by the performance of all systems and subsystems that collectively deliver the functional effects of the ship.

As such, ship trustworthiness is considered to be an aggregation of the trustworthiness of all sub-systems that combine to deliver its performance. It is important to note that different tasks may require trust in different systems (or combinations of systems), and that the way in which individual system trust is assessed may differ from system to system across the ship.

This section describes the way in which trustworthiness is assessed across the sub-systems of a ship, and aggregated together to produce a single measure of ship trustworthiness.

 http://doi.org/10.24868/issn.2515-8198.2019.002

3.1. Functional Ship Breakdown

3.1.1. Key Concepts

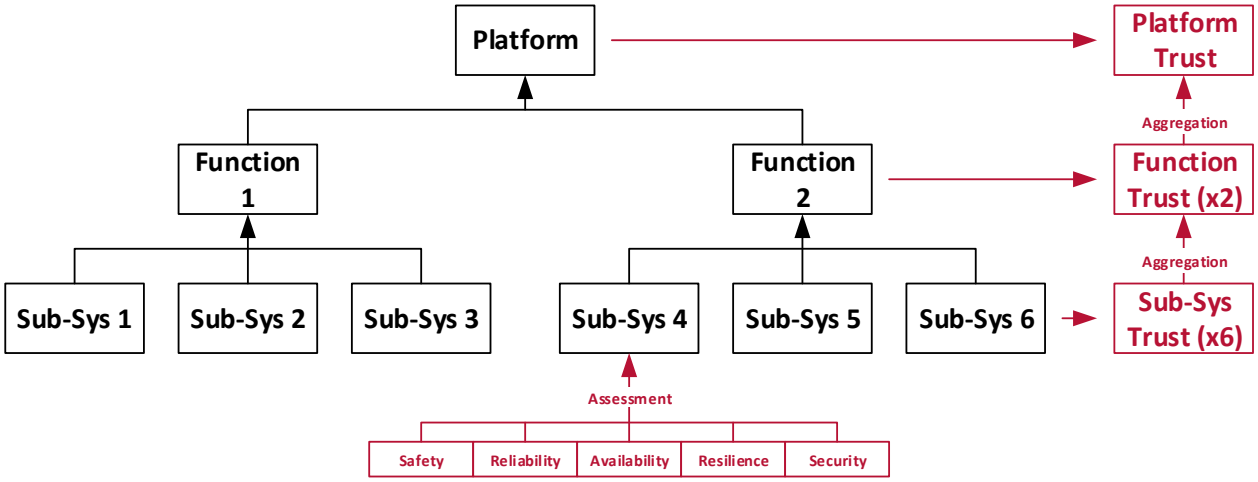

Figure 1: Functional Ship Decomposition, showing how trustwortihness can be calculated at sub-system level and aggregated up, via ship functions, to a single ship trustworthiness assessment

With reference to Figure 1, the trustworthiness assessment process starts at sub-system level, with performance assessed for each of the trustworthiness factors identified in Table 1. A sub-system in this instance is considered to be a complete equipment item, for example a diesel generator, a gearbox, or a radar.

Just as the trustworthiness framework is broken down into several factors that individually impact the ship's ability to function, so too is the ship itself broken down into functional groupings of systems that each provide a different aspect of functionality.

Considering Figure 1, it can be seen for example that the ship is decomposed into a number of 'Function' groups, each of which is delivered by a number of sub-systems; for example, 'Function 2' is provided collectively by Sub-Systems 4, 5, and 6. This may represent 'Propulsion' being provided by a motor and a gearbox, or 'Power Management' being provided by a generator, a switchboard, a power management controller, and cabling. Using these sub-system to function mappings, trustworthiness can be aggregated up to function level as a measure of the ship's ability to deliver that function given the trustworthiness of the sub-systems that provides it.

Finally, the trustworthiness assessed for each function can be aggregated further into a single measure of ship trustworthiness given the trust placed in its ability to deliver the functions that collectively define the ship's capability.

For the purpose of trustworthiness calculation, trust is therefore calculated at the lowest level (in this case the individual sub-systems themselves), aggregated up to provide a measure of trust in the ability of the ship to deliver each function group, and finally aggregated again to provide a measure of trust in the ability to operate as a whole.

The output from this trustworthiness process is not solely applicable at the ship level though, rather at each stage in the process. For example, an equipment operator could monitor the trustworthiness associated with a sub-system as part of their routine operating procedures in an effort to detect signs of degraded performance. Alternatively, a captain may wish to understand the level of trust placed in the propulsion function before undergoing a complex manoeuvre where the communications system trustworthiness is less important. And finally, a remote operator may wish to understand the whole-ship trustworthiness before handing over full autonomous control to the ship to ensure that there are no trustworthiness issues in any function offered by the ship.

3.1.2. Functional Decomposition

The process by which the ship functional decomposition is undertaken is conducted at a high-level such that the method used and resultant functional sub-systems are generically applicable to all ships:

 http://doi.org/10.24868/issn.2515-8198.2019.002

Sub-System	Description
Propulsion	Systems whose primary function is to propel the ship, or to provide control or ancillary functionality in support of this purpose.
Vehicle Control	Systems whose primary function is to control the motion of the ship, or to provide control or ancillary functionality in support of this purpose.
Sensing	Systems whose primary function is the direct interpretation of the external environment.
Power Management	Systems whose primary function is the generation, control, and distribution of power across the ship.
Communications	Systems whose primary function is the transfer of data within and between ships.
Structure	Systems whose primary function is the physical support and containment of other systems.
Mission Planning	Systems whose primary function is the determination of routes, timings and activities to be conducted by the ship.

Table 3: Ship Functional Breakdown (for Illustrative Purposes)

Splitting a ship into functional groups provides a set of perspectives or contexts for considering the effect of an issue being identified. Each functional group is distinct, though the nature of control and communication feedback means that any effect at the ship level is a complex combination of functional effects depending on these interactions. This approach is modular, allowing functions to be added or removed as applicable for a specific ship.

3.2. Trustworthiness Assessment

The trustworthiness level of each sub-system is assessed independently before being aggregated to form a functional group and subsequently a whole-ship trustworthiness level. In the self-assessment of trust, there are two primary factors that a system must consider:

- Inherent restrictions to trust based on the design, operational history, or maintenance. These shall be referred to as static factors, as they will not vary during an operation.
- Live detection of anomalous behaviour indicating a current effect on function. These shall be referred to as dynamic factors, and may vary during operation.

3.2.1. Static Factors

Static factors are calculated at a sub-system level and can act as an upper threshold or scaling factor to trustworthiness such that operation is limited at its extremes by the inherent strengths and weaknesses of its design and the current state of maintenance; these cannot be overcome during operation. Static factors can be further split into those factors which change with time, including historical maintenance or performance metrics, and those factors fixed by design.

Ultimately, each sub-system within a ship must be assigned a single trustworthiness level threshold value representing the upper allowable level of trust that can be placed in a system in operation. Using the judgement of suitably qualified and experienced individuals, systems are assed for each of the three static factor criteria – Design, Operational History, and Maintenance – against the five trustworthiness factors. These static assessments must be performed prior to a mission, as they may change over a ship's operational life

The output of this assessment shall be captured, with justification where required, in matrix form:

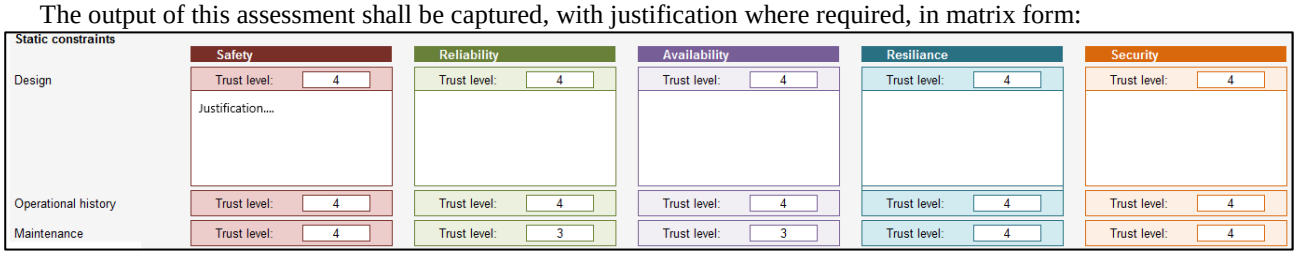

Figure 2: Example sub-system static factor trustworthiness threshold determination

In this example there are no concerns with the Design or Operational History aspects of the system in question, as indicated by 4s being selected across the board for each Aspect. However, maintenance issues identified through periodic inspection mean that Trustworthiness of the mission planning system is reduced to 3 for both the Reliability and Availability aspects.

 http://doi.org/10.24868/issn.2515-8198.2019.002

The output of this process sees the lowest value present, in this case a trustworthiness level of 3, carried forward as the limiting threshold value for the allowable trust that can be placed in this system.

3.2.2. Dynamic Factors

Dynamic trustworthiness assessment characterises the trustworthiness of a ship during operation, through prior understanding of the trustworthiness implications of a range of anomalous events that may occur.

In order to provide context for the effect of an anomaly being detected in a system, a number of signals may be monitored during operation and are grouped into three types as described in Table 4. This grouping provides context for the consideration of the effect of an issue which might affect trustworthiness.

Signal Type	**Description**
Performance	These signals monitor the current output as compared with the requested or typical output. It is therefore a measure of how well the system is currently delivering the function request of it. Examples could be the electrical output of a generator (for example its voltage level or quality-of-supply), or the signal-to-noise ratio of a sensor or transducer.
Capability	These signals monitor the available output compared with the maximum available. It is a measure of the overhead level of performance available in case of a fault or other issue. Examples could be the number of operational generators or the number (or variety) of alternative sensors measuring a given parameter.
Health	These signals monitor the current health status of the system. It is a measure of the current and future potential to deliver the requested function. Examples could be vibration loads on a turbine or error rate on a communications link.

Table 4: Signal Type Description

Importantly, the assessment of dynamic factors against these categories shall be conducted on the ‘real world’ values associated with their operation rather than the measured voltages or currents associated with their detection and measurement. For example, this would see ship speed considered in knots rather than a 4-20mA sensor signal, or switchboard voltage considered as the voltage itself rather than as transducer feedback. This shall allow for the trustworthiness assessment logic to be more easily understood, and more easily applied to a range of ships through the like-for-like comparison of operational properties, without the need to remake or reconfigure the logic to handle different types of sensor etc.

It also reduces the signal monitor requirements to consider only those signals that capture a measure of performance applicable to the function delivered. For example, a generator need only be monitored for its electrical output rather than its many other measurable properties.

3.2.3. Anomaly Detection & Characterisation

Anomaly detection algorithms are capable of characterising a signal and monitoring this signal in real-time to identify anomalous behaviour. This can be performed on any type of signal due to the ship and protocol agnostic nature of these algorithms. This in turn means that the significance, or impact, of an anomaly being detected on a signal is directly affected by what is selected to be monitored, placing emphasis on the appropriateness of signal selection for monitoring. This selection should ensure broad system coverage and be at a consistent level to have meaningful output.

Anomalous behaviour can be detected by deviation from expected behaviour (based on previous experience or design), or by heuristic rules defined using prior knowledge. The combination of these two approaches provides a rounded approach to monitoring considering the current state of a system as well as its design limits.

Anomalies detected during monitoring can be grouped into types based on the trigger for flagging as anomalous. For example, a fault within a gas turbine may be detected as a shaft speed signal exceeding a predetermined limit flagging a ‘Hard Limit’ anomaly.

The five types of anomalies identified are described in Table 5. The combination of type of anomaly detected, nature of monitoring signal, and the source system in question shall collectively indicate the significance of the issue and its effect on the confidence in the system function.

 http://doi.org/10.24868/issn.2515-8198.2019.002

Anomaly Type	Description
Total Loss	Loss of monitoring signal, indicating that monitoring is no longer able to provide information on the current state of the system. State of the system is unknown.
Hard Limit	Monitoring signal exceeds a specified hard limit set to indicate that the system has entered a state which affects its function. State of system is known to significantly affect its function.
Soft Limit	Monitoring signal exceeds a soft limit set to indicate that the system function may be affected, or indicate the onset of an issue. State of system is known to affect its function.
Increased Variance	Variability, or noise, in monitoring signal increased, potentially indicating stability issues in the system or its monitoring and/or the onset of an issue. Stability of system state is affected.
Value Drift	Typical base level of signal is changing, potentially indicating the onset of an issue or system degradation. State of system may degrade function.

Table 5: Monitoring Signal Anomaly Type Definitions

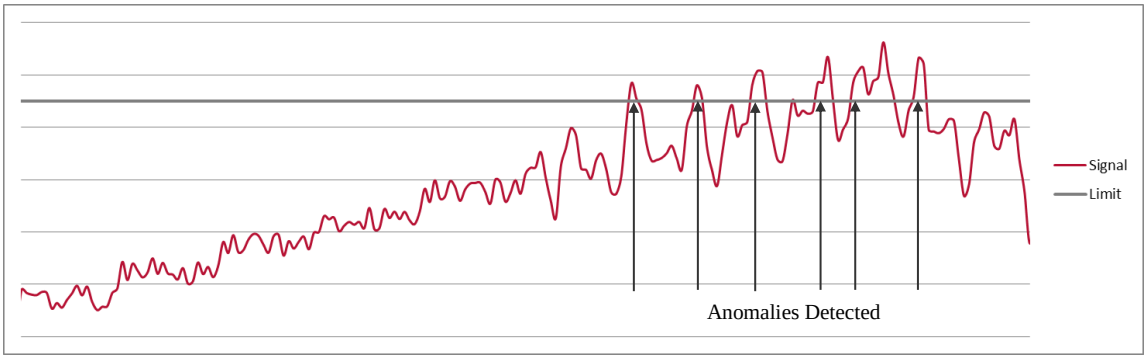

Figure 3: Hard Limit Anomaly Detection

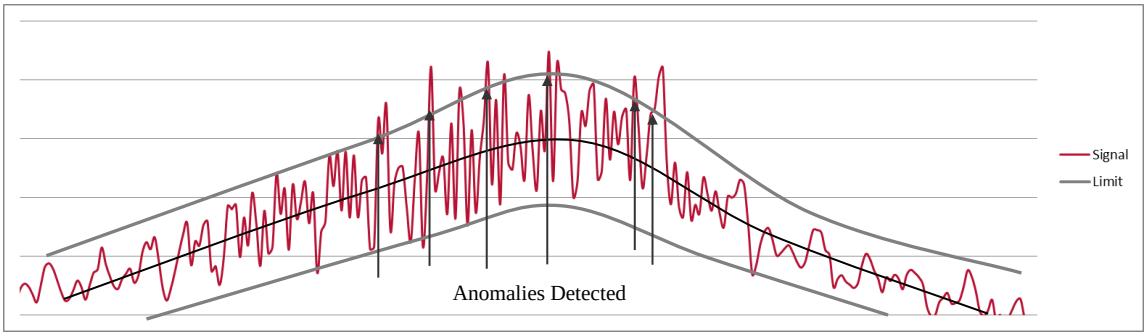

Figure 4: Increased Variance Anomaly Detection

3.3. Assessment and Aggregation

To provide a means of bringing together the results of the static factor assessment conducted prior to operation, and the dynamic factor assessments conducted during operation through anomaly detection, a set of trustworthiness matrices are generated through subject matter expert review to form reference values for use during operation. Assessment of the effects of identifying anomalous behaviour on Trust before operation enables determination of the trustworthiness level of an unmanned ship to be quick and require minimal on-board computation.

3.3.1. Assessment

In order to dynamically assess the trustworthiness of a system during operation a number of signals may be monitored. The trust implications of an anomaly occurring within a sub-system will vary depending on the type of system and the nature of the signal.

 http://doi.org/10.24868/issn.2515-8198.2019.002

Through creating a mapping between anomaly types and trustworthiness aspects, a reference matrix is created for each sub-system which provides a set of trustworthiness levels for a given anomaly type occurrence. This process is performed for each signal type (see Table 4). The result of this is a set of trustworthiness matrices that can be condensed through trustworthiness aggregation into trustworthiness vectors.

Table 6 gives an example trustworthiness matrix for a performance signal type in the propulsion motor sub-system. The example shows that it is considered that the presence of an increased variance type anomaly indicates medium trust in its ability to perform safely. Likewise, that its presence will not affect trust in its ability to perform tasks when requested (i.e. its Availability).

Anomaly Type	Safety	Reliability	Availability	Resilience	Security
Total Loss	1	1	3	3	1
Hard Limit	2	1	3	3	4
Soft Limit	3	3	4	3	4
Increased Variance	3	2	4	4	3
Value Drift	3	3	4	3	2

Table 6: Example Anomaly Type-to-Signal Category Trustworthiness Matrix

3.3.2. Aggregation Algorithm

The trustworthiness outcome of an event occurring can be calculated by Equation 1, where a_i are the trustworthiness levels of each aspect (see Table 5), n is the number of aspects, and *floor* rounds down to the nearest integer value (erring on the side of 'caution' rather than 'optimism'):

$$Trustworthiness\ Level = floor\left(\frac{\sum_{i=1}^{n} a_i}{n}\right)$$

Equation 1: Trustworthiness Aggregation

Aggregated values using Equation 1 are displayed as the trustworthiness vector (final column of Table 7) as a demonstration of this approach.

	Safety	Reliability	Availability	Resilience	Security	Mean	TL
Total Loss	1	1	3	3	1	1.8	1
Hard Limit	2	1	3	3	4	2.6	2
Soft Limit	3	3	4	3	4	3.4	3
Increased Variance	3	2	4	4	3	3.2	3
Value Drift	3	3	4	3	2	3.0	3

Table 7: Example Aggregation of Trustworthiness Matrix to Trustworthiness Vector

In the above example shown in Table 7, it can therefore be seen that for the sub-system in question, the presence of an 'increased variance' anomaly will lead to the system being assessed a trustworthiness level of 3 (unless limited through a static factor pre-assessment).

Ship level trustworthiness and function level trustworthiness in turn are both determined by taking the minimum trustworthiness vector of the components one level below them, as illustrated in Figure 5.

This is only one possible means of aggregation that could be employed; the most suitable algorithm in practice shall depend on a number of factors, such the number of inputs and the operating philosophy for a given sub-system or ship. The philosophy itself could be variable based on the ship, its operation, or its tasking. They may even be a number of aggregation schemes available as required by different mission or operations.

Operators have control to override any anomalies detected prior to aggregating to form a sub-system trustworthiness level. This ensures that sub-system (and consequently ship and function) trustworthiness determination is not unduly influenced by trustworthiness changes in sub-systems that are not required for a particular operation.

 http://doi.org/10.24868/issn.2515-8198.2019.002

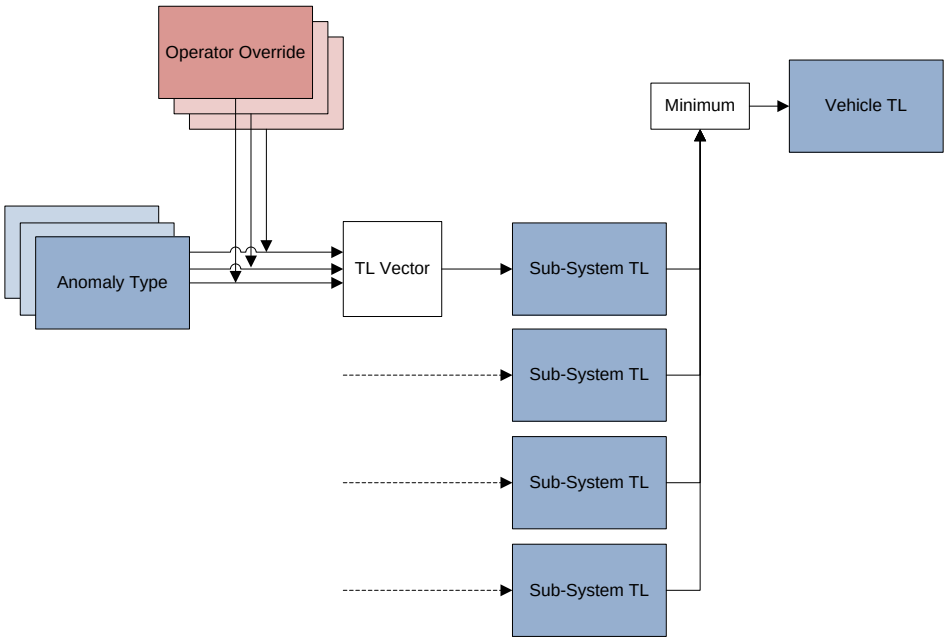

Figure 5: Sub-System and Ship Aggregation System

3.3.3. Trustworthiness Assessment & Aggregation Summary

Figure 6 presents a summary of the architecture through which trustworthiness can be assessed as a factor of: the functional decomposition of the ship; the monitoring of operational signals associated with equipment sub-systems and the detection of anomalies on these signals; pre-existing design or maintenance driven performance factors that operate as a hard limit on trust; and operator interaction in the overriding of not-applicable trustworthiness factors.

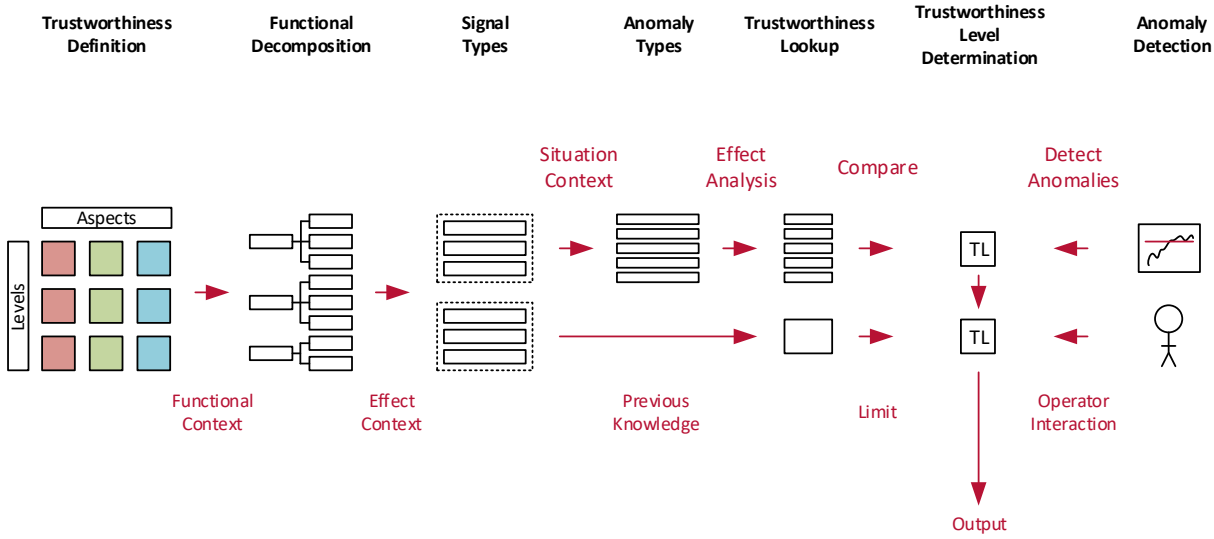

Figure 6: Trustworthiness assessment concept architecture

 http://doi.org/10.24868/issn.2515-8198.2019.002

4. Exploitation

To demonstrate how the concepts presented in this paper can be used to address the problems posed in Section 1.1, this sections presents a number of guided examples intended to demonstrate the value that can be added through their deployment in the marine domain.

4.1. Support to Lean-manning Applications

At present, many marine sub-systems require round-the-clock operation, monitoring and observation to ensure their performance remains within tolerance, be that for safety, functional, or any other aspect of operation.

As ships become increasingly lean-manned, there will have to be a shift towards more autonomous sub-systems that can operate with greater levels of independence, and accordingly 'operators' will take up more hands-off monitoring roles rather than hands-on operation roles.

This transition will require operators to oversee a greater number of systems, meaning they will be required to observe a great number of system operating parameters, and to understand a greater number of performance characteristics. In summary, operators will have to monitor, understand, and act upon an increasingly large amount of data, distributed across an increasingly large number of systems.

In recognition of this transition, the trustworthiness concept can support this change through an 'adaptive display' concept: varying the information presented via equipment control and feedback display panels in response to the perceived trust in the system. Rather than have to observe and analyse a number of parameters for a number of systems simultaneously to identify performance anomalies, an operator could have their attention drawn to issues automatically as they arise.

For example, an operating parameter may begin to fluctuate outside of its normal range, not enough to trigger an automatic alarm or trip on the equipment itself, but sufficiently to trigger the anomaly detection of the trustworthiness algorithms. In this case, the operator's display could automatically be reconfigured to display the parameter in question, the impacted equipment, and the impacts of the reduction in trustworthiness given the anomaly seen (i.e. is safety or availability impacted). This would provide the operator with a contextually 'filtered view' of the issue, simultaneously showing more detailed and more relevant information, which would in turn allow them to make a more informed decision more quickly.

Equivalently, the trustworthiness of a given system could dictate the level of autonomy allowed by a system, and thus the amount of control over its operation afforded to an operator. When system trust is high, the system may be given free rein to operate fully autonomously, with only a high-level summary of operation presented to the operator. As trust in the system drops, the operator could be given a focused summary of the types of anomalies observed and their potential impacts, with an increased amount of control passed back to the operator by way of requests for confirmation of actions from the system before enacting. As trust drops further towards a trustworthiness level of zero, full feedback (as for a 'traditional' setup) would be presented, and control would be firmly given back to the operator with no autonomy allowed until the issues are understood and resolved. This highlights the need for operator training to remain in place such that they can respond appropriately as trustworthiness deteriorates, even in more-autonomous applications.

Across the ship, the grouping of systems by function, with a single operator given responsibility for monitoring (and operating as required) will significantly reduce the manning requirements for the ship. Furthermore, the use of system trust as a means of dictating the information presented to, and the level of control requested from, an operator will allow for increased 'situational awareness' of operators whilst allowing fewer ship staff to oversee an increased number of systems.

From a ship perspective, an understanding of where trust lies in terms of the ship's ability to deliver key functions will prove invaluable to ships' captains when determining and enacting a course of action. For example, a captain that can see that ship trust is low in the ship's 'Propulsion' capability may choose to increase his distance to nearby vessels or landmarks. The trustworthiness concept again provides the captain with increased situational awareness through aggregating low-level anomalies into a more useful parameter that presents the impact of those issues on their ability to deliver a ship level capability.

4.2. Support to More-Autonomous Ships

To combat the increase in shipping densities seen globally, combined with the usage of more- and fully-autonomous ships, shipping operators and the captains of individual ships alike are going to have to adapt to different ways of operating, without sacrificing the safety of their assets, those around them, and the environment.

The trustworthiness concept provides a way of adaptively varying the allowable proximity of vessels to each other or to landmarks through 'exclusion zones' given the trust placed by each vessel in its ability to operate at a given point in time:

 http://doi.org/10.24868/issn.2515-8198.2019.002

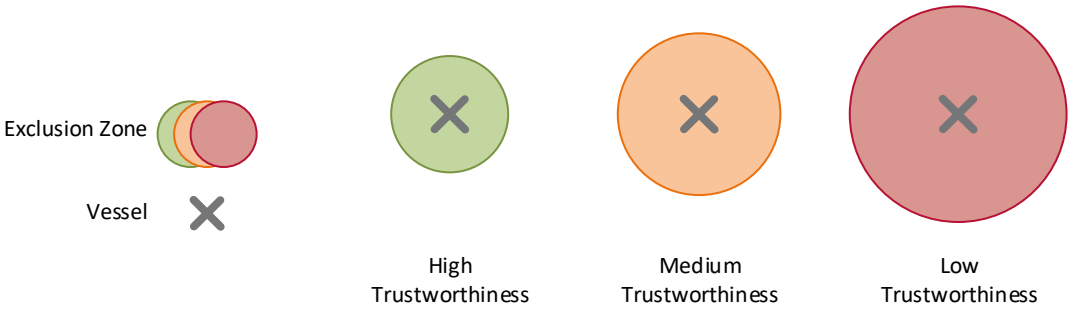

Figure 7: Using ship trustworthiness to vary the size of a ship's exclusion zone

This concept provides ships with a method through which they can operate safely in closer proximity to one another while trust in the ship, and therefore trust in its ability to operate safely and when demanded, is high. As trust in a given ship drops however and the exclusion zone increases, this degradation is seen by and shared with nearby vessels who can respond accordingly, either through the ship in distress removing itself from theatre, or by nearby vessels varying their course accordingly..

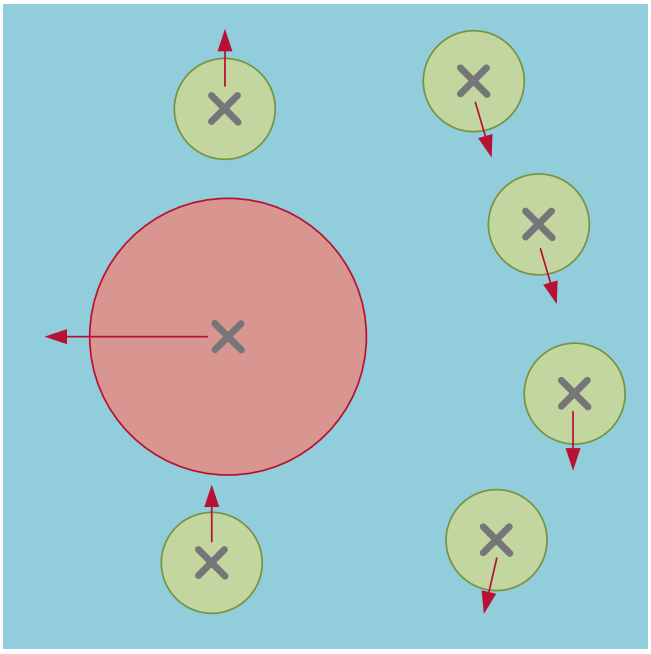

Figure 8: Adaptive operating proximity based on the self-assessed trust of each vessel

Considering the example shown in Figure 8, which shows a number of vessels operating in a shipping lane, one vessel has assessed a reduced level of trust and so has increased its safe operating area accordingly. In response to this change, the vessels around it have given it a wider berth to minimise the risk of collision, and the vessel in trouble has set an evasive course to remove itself from harm's way.

In a fully-autonomous future, this could all happen automatically, with trustworthiness assessments continually being conducted and relayed between vessels to allow for adaptive mission planning and routing to take place. As an interim step the trustworthiness assessments would provide captains with a better understanding of where they position themselves relative to manned and unmanned vessels to minimise the risks posed to their ship. As such, the utilisation of this concept is equally applicable to both manned and autonomous vessels.

 http://doi.org/10.24868/issn.2515-8198.2019.002

5. Conclusions

This report presents a framework that defines the concept of 'trustworthiness' in marine ships. A set of five aspects for trust is presented, refined from an industry accepted framework of trust, that fully define trustworthiness in this context. Assessment by subject matter experts form references for the in-operation assessment of ship trustworthiness, and a scale of trustworthiness levels is presented, drawing upon similar concepts seen in well-known SILs and DALs as applied to safety assessment.

The concept has been developed considering ship and system level perspectives, though the frameworks and processes have been designed to be generic, modular and readily extensible as required. Wider applicability and scalability of these outputs would form a logical step for future investigation. The concept is presented as a demonstrable baseline and starting point for refinement given considerations of practicability, extensibility, and scalability.

Example applications describing a method of providing additional contextual information to equipment/sub-system operators and ship captains alike has been investigated, in addition to the way in which the concept is applicable to both manned, lean-manned, and fully autonomous ships. Benefits resulting from these applications include clearer and more coherent information access and understanding, more focused operator workload, and improved and quicker situational awareness.

The concept has been developed with potential for refinement through further input from experts and testing of its extensibility, both in terms of breadth (across different ships and applications) and depth (increased or decreased level of system fidelity). Such further work would mature the concept and develop it for inclusion throughout the lifecycle of a ship as a core embedded capability, rather than a later-stage bolt-on capability.

The concept of trustworthiness of unmanned ships is applied in a structured way which is informed both by subject matter experts (considering ship-specific effects of issues on trustworthiness) and real-time events (identification and categorisation of issues through signal monitoring). Trustworthiness levels are a way of quantifying the level of confidence in a system to function as expected and communicating this to the control authority, and are generically applicable to sub-systems and ships alike. The framework allows for a more informed decision to be made, with known confidence, for how to respond to a scenario based on the degradation in trust seen and what the ship-level impacts are.

The application of an automated trustworthiness framework is novel in this context and could have a significant impact upon the future design and deployment of unmanned marine ships, as well as broader applicability to autonomous and unmanned ships in other domains.

References

[1] PAS 754: 2014, Software Trustworthiness – Governance and management – Specification. Trustworthy Software Initiative, 2014.

[2] FNC47206-42823R, Self Aware Trustworthiness Levels and Information Dissemination (SATLID). Frazer-Nash Consultancy, 2015.

[3] FNC49048-45595R, Self Aware Trustworthiness Level and Assurance with Operational Policy (SATLAOP). Frazer-Nash Consultancy, 2017.

 http://doi.org/10.24868/issn.2515-8198.2019.002