./Figures/logoc.pdf

An Investigation into Trust and Reputation Frameworks for Autonomous Underwater Vehicles

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy by

Andrew Bolster

June 2015

Contents

Notations **vii**

Preface **ix**

Abstract **xi**

Acknowledgements **xiii**

1 Introduction **1**

2 Background on Trust and its Applications to MANETs **3**

2.1 Trust 3

2.2 Trust in MANETs 4

2.2.1 Design Considerations 6

2.2.2 Current Trust Management Frameworks 7

2.2.3 Trust as an incomplete system characteristic 8

2.3 Grey System Theory and Grey Trust Assessment 9

2.3.1 Grey numbers, operators and terminology 9

2.3.2 Whitenisation and the Grey Core 10

2.3.3 Grey Sequence Buffers and Generators 10

2.3.4 Grey Trust 11

2.3.5 PROSE: Whats the point 12

3 Maritime Communications Environment and Use of Autonomous Systems **15**

3.0.6 Trust in Marine Networks 16

4 Trust in Autonomous Systems of Systems for Maritime Defence Applications **17**

4.1 Trust Perspectives 19

4.1.1 Design Trust 21

Current Unmanned System Interface Standardisation 22

NATO Standardization Office 23

Society of Automotive Engineers (SAE) 24

American Society of Testing and Materials (ASTM) 24

4.1.2 Operational Trust 25

Information Overload 25
Adaptive Automation 25
Distributed Decision Making 26
Complexity 26
Cognitive Biases and Failing Heuristics 27
Summary of Human Factors impacting Operational Trust in Defence Contexts 28
4.2 Trust and Reputation in Autonomous Collaborative Systems 28
4.3 Levels of Trust 29

5 Strategies for Multi-Domain Trust Assessment **31**

6 Modelling and Analysis of Collaborative Node Kinematic Behaviuors in Underwater Acoustic MANETs **33**
6.0.1 Establishing Scale Factors in Communications Rate 33
6.0.2 Establishing Scale Factors in Physical Distribution 34
6.0.3 Metric Weighting 38

7 Comparative Analysis of Multi-Domain Trust Assessment in Collaborative Marine MANETs **41**

Bibliography **43**

Illustrations

List of Figures

6.1 Varying packet emission rate demonstrates maximal throughput at 0.025 packets per second, equivalent to ≈240 bps . . . 34
6.2 Varying packet emission rate demonstrates a saturation point at 0.025 packets per second . . . 34
6.3 Comparison of Medium Acquisition Collisions, Throughput, and Enqueued packets against varying application packet emission rates. . . . 35
6.4 Probability of Timely Reception across a range of node scaling. . . . 36
6.5 End to End Delay under varying node-separations . . . 37
6.6 RTS/Data ratio for varying node-separations . . . 37
6.7 MTFM Trust assessments for varying mobility options in the selfish case . . 39
6.8 Beta Trust time varying assessments for of $n1$ varying mobility options . . . 40

List of Tables

2.1 Comparison between selected methods of characterising uncertainty, adapted from [5] [9] [13] [16] . . . 9

4.1 Examples of Roles that require a Design Perspective of Trust in Autonomous Systems. . . . 20
4.2 Examples of Roles that require a Operational Perspective of Trust in Autonomous Systems. . . . 21
4.3 Levels of Interoperability for STANAG 4586 Compliant UCS . . . 23

6.1 Tabular view of data from Figs 6.4, 6.5, and 6.6 . . . 38

Notations

The following notations and abbreviations are found throughout this thesis:

Preface

This thesis is primarily my own work. The sources of other materials are identifed.

Abstract

As Autonomous underwater vehicles (AUVs) become technically more competent, and fiscally more attainable, their use has been applied to a great many areas within defence, commercial and environmental areas of concern. Increasingly,these applications are tending towards utilising independent collective behaviour of teams or fleets of these platforms.

Acknowledgements

There are many people who deserve the highest thanks for their support, patience, kindness and understanding. The greatest thanks have to be distributed among my family and friends, for putting up with my madness; both the madness of starting it and the madness of seeing it through. Maybe I'll get a job that you can actually explain! Next, I must thank Professor Marshall, without whom this work wouldn't have been attempted let alone completed. Finally, this PhD is dedicated to R, who knows why.

Chapter 1

Introduction

Chapter 2

Background on Trust and its Applications to MANETs

2.1 Trust

In human trust relationships it is recognized that there can be several perspectives of Trust for example organizational, sociological, interpersonal, psychological and neurological [6]. For the purposes of this work we define two perspectives on trust for autonomous systems: Design and Operational. These are summarised as follows:

- *Design Trust*; When an autonomous system is under development a level of Trust is established in it through the manner in which it has been designed and tested. This is the same as conventional systems. The difference with systems that have high-levels of autonomy is that they are designed to behave adaptively to dynamic environments that are difficult to fully predict prior to operational deployment. For example, in a navigation system it is difficult to predict the dynamic environment it will need to adapt to. So Trust needs to be developed that th:e design and test of such systems are sufficient to predict that operation will be, if not optimal, at least satisfactory.

- *Operational Trust*; Trust at runtime or in-situ that both the individual nodes within a system are operating as expected[1]; and that the interfaces between the operator and the system are as expected. This latter aspect covers issues such as physical/wireless links and interpretation of data at each end of such a communication link.

In addition to the two perspectives of trust identified, it is necessary to define and classify Operational Trust into two distinct but related sections, which we define as being:

- *Hard Trust* or technical trust, being the quantitative measurement and communication of the expectation of an actor performing a certain task, based on historic

[1]Operational Trust is functionally derived from, but distinct from Design Trust

performance and through consensus building within a networked system. Can be thought of as a de-risking strategy to measure and monitor the ability of a system, or another actor within a system, to perform a task unsupervised.

- *Soft Trust* or common trust, being the qualitative assessment of the ability of an actor to perform a task or operation consistently and reliably based on social or experiential factors. This is the natural form of trust and is the main motivational driver for the human-factors trust discussion. Can be rephrased as the level of confidence an operator has in an actor to perform a task unsupervised.

It is already clear that these two definitions are extremely close in their construction, but represent fundamentally different approaches to trust, one coming from a sociological perspective of person-to-person and person-to-group relationships from day to day life, and the other coming from a statistical or formal appraisal of an activity by a system. For the purposes of this work, we are concerned with the analytical establishment of hard trust within a topologically dynamic network of autonomous actors.

2.2 Trust in MANETs

As mobile ad-hoc networks (MANETs) grow beyond the terrestrial arena, their operation and the protocols designed around them must be reviewed to assess their suitability to different communications environments, ensuring their continued security, reliability, and performance.

Trust Management Frameworks (TMFs) provide information to assist the estimation of future states and actions of nodes within networks. This information is used to optimize the performance of a network against malicious, selfish, or defective misbehaviour by one or more nodes. Previous research has established the advantages of implementing TMFs in 802.11 based MANETs, particularly in terms of preventing selfish operation in collaborative systems [7], and maintaining throughput in the presence of malicious actors [2]

Most current TMFs use a single type of observed action to derive trust values, i.e. successfully forwarded packets. These observations then inform future decisions of individual nodes, for example, route selection [8].

Recent work has demonstrated use of a number of metrics to form a “vector” of trust. The Multi-parameter Trust Framework for MANETs (MTFM)[5], uses a range of physical metrics beyond packet delivery/loss rate (PLR) to form a vector of trust. This vectorized trust allows a system to detect and identify the tactics being used to undermine or subvert trust. To date this work has been limited to terrestrial, RF based networks, however as autonomous underwater vehicles (AUVs) become more capable, and economical, they are being used in many applications requiring trust. These applications are using the collective behaviour of teams or fleets of these AUVs to accomplish tasks [3]. With this use being increasingly isolated from stable communications networks, the establishment of trust between nodes is essential for the reliability and stability of such teams. As such, the use of trust methods developed in the terrestrial MANET space must be re-appraised for application within the challenging underwater communications channel.

The distributed and dynamic nature of MANETs mean that it is difficult to maintain a trusted third party (TTP) or evidence based trust system such as Certificate Authorities (CA) or Public Key Infrastructure (PKI). Distributed trust management frameworks aim to detect, identify, and mitigate the impacts of malicious actors by distributing per-node assessments and opinions to collectively self-police behaviour. Various models and algorithms for describing trust and developing trust management in distributed systems, P2P communities or wireless networks have been considered. Taking some examples;

- *The Objective Trust Management Framework* takes a Bayesian Beta function to model per-link Packet Loss Rate (PLR) over time, combining "Trust" and "Confidence of Assessment" into a single value [8]. OTMF however does not appropriately combat multi-node-collusion in the network [4].
- *Trust-based Secure Routing*[12] demonstrated an extension to Dynamic Source Routing (DSR), incorporating a Hidden Markov Model of next-hop network, reducing the efficacy of Byzantine attacks such as black-hole routing.
- *CONFIDANT*[2] presented an approach using a probabilistic estimation of PLR, similar to OTMF, also introducing a topology weighting scheme that also weighted trust assessments based on historical experience of the reporter.
- *Fuzzy Trust-Based Filtering*; [11] presents the use of Fuzzy Inference to adapt to malicious recommenders using conditional similarity to classify performance with overlapping Fuzzy Set Membership, filtering assessments across a network.

These TMFs can be generalised as single-value probabilistic estimation, based around using a binary input state and generating an probabilistic estimation of the future states of that input. This expectation value is $\mathrm{beta}(p|\alpha, \beta) \rightarrow E(p) = \frac{\alpha}{\alpha+\beta}$ where α and β represent the number of successful and unsuccessful interactions respectively.

These single metric TMFs provide malicious actors with a significant advantage if their activity is undetectable by that metric. In the case where the attacker can subvert the TMF, the metric under assessment by that TMF does not cover the threat mounted by the attacker. In turn, this causes a super-linearly negative effect in the efficiency of the network, as the TMF is assumed to have reduced the possible set of attacks when it has actually made it more advantageous to attack a different part of the networks operation. An example of such a situation would be in a TMF focused on PLR where an attacker selectively delays packets going through it, reducing overall throughput but not dropping any packets. Such behaviour would not be detected by the TMF.

There are also situations where the observed metrics will include significant noise and occur at irregular, sparse, intervals. Conventional approaches such as probabilistic estimation do not produce trust values that reflect the underlying reality and context of the metrics available, as they require a-priori assumption that the trust value under exploration has an expected distribution, that distribution is mono-modal, and the input metrics are binary. In scenarios with variable, sparse, noisy metrics, estimating the distribution is difficult to accomplish a-priori.

2.2.1 Design Considerations

There are five topics that are important to address in any MANETs trust model [**?**]:

1. The trust model should be without infrastructure. Because the network routing infrastructure is formed in an ad-hoc fashion, the trust management can not depend on, e.g., a trusted third party (TTP). There is no public key infrastructure (PKI), where some center nodes monitor the network, and publish illegal nodes periodically. In a MANET, there are no certification authorities (CA) or registration authorities (RA) with elevated privileges etc.

2. The trust model should be anonymous because of the anonymity of mobile nodes in MANETs.

3. The trust model should be robust. That is, it can be robust to all kinds of unfriendly attacks and the network itself should not be susceptible to attacks by unfriendly nodes. Moreover, in the presence of malicious nodes, they attempt to subvert the model in order to get the unfairly good trust value.

4. The trust model should have minimal control overhead in accordance with computation, storage, and complexity.

5. The trust model should be self-organized. MANETs are characterized to have dynamic, random, rapidly changing and multi-hop topologies composed of relatively bandwidth-constrained

Trust is the level of confidence one agent has in another to perform a given action on request or in a certain context. Trust in the autonomous or semi-autonomous realm is the ability of a system to establish and maintain confidence in itself or another systems' operations. Managing this trust can be used to predict and reason on the future interactions between entities in a system, such as an autonomous mobile ad-hoc network (MANET).

The distributed and dynamic nature of MANETs mean that it is difficult to maintain a trusted third party (TTP) or evidence based trust system such as Certificate Authorities or using Public Key Infrastructures (PKI). Therefore, a distributed, collaborative system must be applied to these networks. Such distributed trust management frameworks aim to detect, identify, and mitigate the impacts of malicious actors by distributing per-node assessments and opinions to collectively self-police behaviour.

2.2.2 Current Trust Management Frameworks

Various models and algorithms for describing trust and developing trust management in distributed systems, P2P communities or wireless networks have been considered. Taking some examples;

- *The Objective Trust Management Framework* takes a Bayesian approach and introduces the idea of applying a Beta function to changes in the per-link Packet Loss Rate (PLR) over time, combining “Trust” and “Confidence of Assessment” into a single value [8]. OTMF however does not appropriately combat multi-node-collusion in the network [4].
- *Trust-based Secure Routing [12]* demonstrated an extension to Dynamic Source Routing (DSR), incorporating a Hidden Markov Model of the wider ad-hoc network, reducing the efficacy of Byzantine attacks, particularly black-hole attacks but is limited by focusing on single metric observation (PLR)[4].
- *CONFIDANT*; [2] presented an approach using a probabilistic estimation of normal observations, similar to OTMF. They also introduced a greedy topology weighting scheme that internally weighted incoming trust assessments based on historical experience of the reporter.
- *Fuzzy Trust-Based Filtering*; [11] presented a method using Fuzzy Inference to cope with imperfect or malicious recommendation based on a probabilistic estimation of performance using conditional similarity to classify performance using overlapping Fuzzy Set Membership functions to collaboratively filter reputations across a network.

OTMF, CONFIDANT, and Fuzzy Trust-Based Filtering can be generalised as single-value probabilistic estimation, based around a Bayesian idea of taking a binary input state and generating an idealised Beta Distribution (2.1) of the future states of that input generated through an expectation value based on interactions (2.2).

$$\text{beta}(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}, \text{ where } 0 \leq p \leq 1; \alpha, \beta > 0 \tag{2.1}$$

$$E(p) = \frac{\alpha}{\alpha+\beta} \tag{2.2}$$

Where α and β represent the number of successful and unsuccessful interactions respectively.

These single metric TMFs provide malicious actors with a significant advantage if their activity is undetectable by that one assessed metric, especially if the attacker knows the metric in advance.

The objective of operating a TMF is to increase the confidence in, and efficiency of, a system by reducing the amount of undetectable negative operations an attacker can perform. In the case where the attacker can subvert the TMF, the metric under assessment by that TMF does not cover the threat mounted by the attacker. In turn, this causes a super-linearly negative effect in the efficiency of the network as the TMF is assumed to have reduced the possible set of attacks when in fact it has only made it more advantageous to attack a different aspect of the networks operation. An example of such a behaviour would be the case in a TMF focused on PLR where an attacker selectively delays packets going through it, reducing the over all throughput of one or more virtual network routes. Such behaviour would not be detected by the TMF.

Many trust systems operate on the basis of establishing closed system models based on noisy or perturbed information inputs, sourced by decentralised agents or nodes, with an aim to collaboratively establishing additional information about the expected states and behaviours of other agents within a system. As such, trust systems can be described as fundamentally uncertain, particularly in the areas or reputation establishment and trust chaining.[**?**]. Adding to this state the highly dynamic features of many aspects of trust theory applications (Ad Hoc Networks, Online Markets, etc.), we can generalise the sources of incomplete information from a single nodes perspective as being part of 4 cases.

say some- hat 'agent' de' are used ngeably in ument

- Information on the system's boundary is incomplete
- Information about the range of system behaviours is incomplete
- Information about the structure of the system is incomplete or out of date
- Information about observed parameters (metrics) is incomplete or out of date.

These cases of incompleteness of information are closely mirrored by those for which grey theory was originally posited as a form of system modeling, putting information incompleteness at the centre of the assessment. While some work [5] has been done to apply grey theory to a trust context, it has not been fully explored. Guo applies grey analysis to generate a "trust vector" from the grey whitenisation of independent or near-independent metrics. In this paper we demonstrate a methodology that applies Grey Sequence operations and Grey Generators (conceptually analogous to Sequential Bayesian Filtering") to provide continuous trust assessment in a sparse, asynchronous metric space across multiple domains of trust.

2.2.3 Trust as an incomplete system characteristic

While application specific trust management frameworks are often based on a very limited space of available metrics, the problem of establishing trust in dynamical systems such as social, economic or autonomous systems have the opportunity to tap in to a wide range of potential metric spaces. Taking the example of Mobile Ad-Hoc Networks (MANET), the variable most applied to the assessment of trust is the packet error rate, or more generally, the number of successful and unsuccessful interactions between two agents within a system. However, a wealth of other information is available within this example; for instance the delay in communications from one node to another; the total throughput of particular network links; and in the case of wireless networks, the strength of received signals. Looking beyond the communications domain, within such a MANET, information is also usually available regarding the physical domain of a network; the relative positioning and motions of nodes within a network can also be used to inform the generation of trust assessments.

Table 2.1 provides a qualitative summary of the differences in use and application between Fuzzy, Probabilistic and Grey Systems of managing uncertainty.

TABLE 2.1: Comparison between selected methods of characterising uncertainty, adapted from [5] [9] [13] [16]

	Fuzzy Math	Bayesian Estimation	Grey Systems
Objects	Cognitive Uncertainty	Distribution Refinement	Poor Information
Set Style	Fuzzy Sets	Cantor Sets	Grey Hazy Sets
Processes	Marginal Sampling	Frequency Distribution	Sequence Generation
Requirement	Known Membership	Beta Distribution	Any Distribution
Emphasis	Extension	Intension	Intension
Characteristics	Experience	Large Samples	Small Samples

2.3 Grey System Theory and Grey Trust Assessment

2.3.1 Grey numbers, operators and terminology

Grey numbers are used to represent values where their discrete value is unknown, where that number may take its possible value within an interval of potential values, generally written using the symbol $\oplus$. Taking a and b as the lower and upper bounds of the grey interval respectively, such that $\oplus \in [a,b] | a < b$ The "field" of $\oplus$ is the value space $[a,b]$. There are several classifications of grey numbers based on the relationships between these bounds.

don't think c
fication is th
word here

Black and White numbers are the extremes of this classification; such that $\dot{\oplus} \in [-\infty,+\infty]$ and $\mathring{\oplus} \in [x,x] | x \in \mathbb{R}$ or $\oplus(x)$ It is clear that white numbers such as $\mathring{\oplus}$ have a field of zero while black numbers have an infinite field.

Grey numbers may represent partial knowledge about a system or metric, and as such can represent half-open concepts, by only defining a single bound; for example $\underline{\oplus} = \oplus(\underline{x}) \in [x,+\infty]$ and $\overline{\oplus} = \oplus(\overline{x}) \in [-\infty,x]$.

Primary operations within this number system are as follows;

$$\oplus_1 + \oplus_2 \in [a_1 + a_2, b_1 + b_2] \tag{2.3a}$$

$$-\oplus \in [-b, -a] \tag{2.3b}$$

$$\oplus_1 - \oplus_2 = \oplus_1 + (-\oplus) \tag{2.3c}$$

$$\begin{aligned}\oplus_1 \times \oplus_2 \in\ &[\min(a_1a_2, a_1b_2, b_1a_2, b_2a_2),\\ &\max(a_1a_2, a_1b_2, b_1a_2, b_2a_2)]\end{aligned} \tag{2.3d}$$

$$\oplus^{-1} \in [b^{-1}, a^{-1}] \tag{2.3e}$$

$$\oplus_1/\oplus_2 = \oplus_1 \times \oplus_2^{-1} \tag{2.3f}$$

$$\oplus \times k \in [ka, kb] \tag{2.3g}$$

$$\oplus^k \in [a^k, b^k] \tag{2.3h}$$

where k is a scalar quantity.

2.3.2 Whitenisation and the Grey Core

The characterisation of grey numbers is based on the encapsulation of information in a grey system in terms of the grey numbers core ($\hat{\oplus}$) and it's degree of greyness (g°). If the distribution of a grey number field is unknown and continuous, $\hat{\oplus} = \frac{a+b}{2}$.

Non-essential grey numbers are those that can be represented by a white number obtained either through experience or particular method. [10] This white hissed value is represented by $\tilde{\oplus}$ or $\oplus(x)$ to represent grey numbers with x as their whitenisation. In some cases depending on the context of application, particular gray numbers may temporarily have no reasonable whitenisation value (for instance, a black number). Such numbers are said to be Essential grey numbers.

2.3.3 Grey Sequence Buffers and Generators

equence
and partial

Given a fully populated value space, sequence buffer operations are used to provide abstractions over the dataspace. These abstractions can be *weakening* or *strengthening*. In the weakening case, these operations perform a level of smoothing on the volatility of a given input space, and strengthening buffers serve to highlight and

A powerful tool in grey system theory is the use of grey incidence factors, comparing the “likeness” of one value against a cohort of values. This usefulness applies particularly well in the case of multi-agent trust networks, where the aim is to detect and identify malicious or maladaptive behaviour, rather than an absolute assessment of “trustworthiness”.

2.3.4 Grey Trust

Grey Theory performs cohort based normalization of metrics at runtime. This creates a more stable contextual assessment of trust, providing a “grade” of trust compared to other observed nodes in that interval, while maintaining the ability to reduce trust values down to a stable assessment range for decision support without requiring every environment entered into to be characterised. Grey assessments are relative in both fairly and unfairly operating networks. Nodes will receive mid-range trust assessments if there are no malicious actors as there is no-one else “bad” to compare against.

Guo[5] demonstrated the ability of Grey Relational Analysis (GRA)[18] to normalise and combine disparate traits of a communications link such as instantaneous throughput, received signal strength, etc. into a Grey Relational Coefficient, or a “trust vector”.

In the case of the terrestrial communications network used in [5], the observed metric set $X = x_1, \ldots, x_M$ representing the measurements taken by each node of its neighbours at least interval, is defined as $X =$ [packet loss rate, signal strength, data rate, delay, throughput]. The trust vector is given as

$$\begin{aligned} \theta^t_{k,j} &= \frac{\min_k |a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|}{|a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|} \\ \phi^t_{k,j} &= \frac{\min_k |a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|}{|a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|} \end{aligned} \tag{2.4}$$

where $a^t_{k,j}$ is the value of a observed metric x_j for a given node k at time t, ρ is a distinguishing coefficient set to 0.5, g and b are respectively the ’“good” and “bad” reference metric sequences from $\{a^t_{k,j} k = 1, 2 \ldots K\}$, e.g. $g_j = \max_k(a^t_{k,j})$, $b_j = \min_k(a^t_{k,j})$ (where each metric is selected to be monotonically positive for trust assessment, e.g. higher throughput is always better).

Weighting can be applied before generating a scalar value which allows the identification and classification of untrustworthy behaviours.

$$[\theta^t_k, \phi^t_k] = \left[\sum_{j=0}^{M} h_j \theta^t_{k,j}, \sum_{j=0}^{M} h_j \phi^t_{k,j} \right] \tag{2.5}$$

Where $H = [h_0 \ldots h_M]$ is a metric weighting vector such that $\sum h_j = 1$, and in the basic case, $H = [\frac{1}{M}, \frac{1}{M} \cdots \frac{1}{M}]$ to treat all metrics evenly. θ and ϕ are then scaled to $[0, 1]$ using the mapping $y = 1.5x - 0.5$. The $[\theta, \phi]$ values are reduced into a scalar trust value by $T^t_k = (1 + (\phi^t_k)^2/(\theta^t_k)^2)^{-1}$. This trust value minimises the uncertainties of belonging to either best (g) or worst (b) sequences in (2.4).

MTFM combines this GRA with a topology-aware weighting scheme(2.6) and a fuzzy whitenization model(2.7). There are three classes of topological trust relationship used; Direct, Recommendation, and Indirect. Where an observing node, n_i, assesses the trust of another, target, node, n_j; the Direct relationship is n_i’s own observations n_j’s behaviour. In the Recommendation case, a node n_k, which shares Direct relationships with both n_i and n_j, gives its assessment of n_j to n_i. The Indirect case, similar to the Recommendation case, the recommender n_k, does not have a direct link with the observer n_i but n_k has a Direct link with the target node, n_j. These relationships give us node sets, N_R and N_I containing the nodes that have recommendation or indirect, relationships to the observing node respectively.

$$\begin{aligned} T_{i,j}^{MTFM} = \frac{1}{2} \cdot \max_s\{f_s(T_{i,j})\}T_{i,j} + \frac{1}{2}\frac{2|N_R|}{2|N_R|+|N_I|}\sum_{n\in N_R}\max_s\{f_s(T_{i,n})\}T_{i,n} \\ + \frac{1}{2}\frac{|N_I|}{2|N_R|+|N_I|}\sum_{n\in N_I}\max_s\{f_s(T_{i,n})\}T_{i,n} \end{aligned} \tag{2.6}$$

Where $T_{i,n}$ is the subjective trust assessment of n_i by n_n, and $f_s = [f_1, f_2, f_3]$ given as:

$$\begin{aligned} f_1(x) &= -x + 1 \\ f_2(x) &= \begin{cases} 2x & \text{if } x \le 0.5 \\ -2x + 2 & \text{if } x > 0.5 \end{cases} \\ f_3(x) &= x \end{aligned} \tag{2.7}$$

2.3.5 PROSE: Whats the point

Grey System Theory, by it's own authors admission, hasn't taken root in it's originally intended area of system modelling [**?**]. However, given it's tentative application to MANET trust, taking a Grey approach on a per metric benefit has qualitative benefits that require investigation; the algebraic approach to uncertainty and the application of "essential and non essential greyness", whiteisation, and particularly grey buffer sequencing allow for the opportunity to generate continuous trust assessments from multiple domains asynchronously;

For a given metric set X such that $X = x_1, \ldots, x_M$ representing the M different types of measurement generated by an observer. If these metrics are not synchronised, for instance if they are interrupt driven such as communications-based observations, generating more abstract measurements requires inherent assumptions about "how to accumulate the data while you wait". For instance, in [1], we demonstrated a periodic trust assessment framework for autonomous marine environments, in such an environment, to establish useful, generalised, data, it was necessary to wait for a relatively long time to accumulate enough data to make assessments. However, this left many 'smells'; data was being left in-buffer for a long time before being used to make decisions, and by the time the data was collated and processed, it could be wildly different from the reality. Further, while some periods could be extremely sparse or even empty, others could be extremely busy with many records having to be averaged down to provide a 'single period' response. Therefore, the implementation of a suitable sequence buffer version of the framework would be beneficial.

Such a sequence buffer framework would involve a tracking predictor that would provide best-guess estimates of an interpolated value for a metric between value updates, and a back-propagation algorithm to retroactively update historical assessments of that metrics so as to better inform any abstracted trust value predictor.

I had initially thought that such a back-propogator would be a total mess as I'd imagined that significant-model-breaking would potetially indicate untrustworthy behaviour, but this is stupid since the per-metric-model has the least information of anyone and is simply there to provide better intermediate values and has no / limited direct impact on the overall trust behaviour.

This backpropogation will probably be a pain to implement as it'd require a retroactive reassessment of trust and could get really messy if it was interrupt driven, but it's better not to prematurly optimise.

Chapter 3

Maritime Communications Environment and Use of Autonomous Systems

The key challenges of underwater acoustic communications are centred around the impact of slow and differential propagation of energy (RF, Optical, Acoustic) through water, and it's interfaces with the seabed / air. The resultant challenges include; long delays due to propagation, significant inter-symbol interference and Doppler spreading, fast and slow fading due to environmental effects (aquatic flora/fauna; surface weather), carrier-frequency dependent signal attenuation, multipath caused by the medium interfaces at the surface and seabed, variations in propagation speed due to depth dependant effects (salinity, temperature, pressure, gaseous concentrations and bubbling), and subsequent refractive spreading and lensing due to that same propagation variation[14].

The attenuation that occurs in an underwater acoustic channel over a distance d for a signal about frequency f in linear and dB forms respectively is given by

$$A_{\text{aco}}(d, f) = A_0 d^k a(f)^d \tag{3.1}$$

$$10 \log A_{\text{aco}}(d, f)/A_0 = k \cdot 10 \log d + d \cdot 10 \log a(f) \tag{3.2}$$

where A_0 is a unit-normalising constant, k is a spreading factor (commonly taken as 1.5), and $a(f)$ is the absorption coefficient, expressed empirically using Thorp's formula (3.3) from [15]

$$10 \log a(f) = 0.11 \cdot \frac{f^2}{1 + f^2} + 44 \cdot \frac{f^2}{4100 + f^2} + 2.75 \times 10^{-4} f^2 + 0.003 \tag{3.3}$$

Refractive lensing and the multipath nature of the medium result in supposedly line of sight propagation being extremely unreliable for estimating distances to targets. The first arriving beam has as the very least bent in the medium, and commonly has reflected off the surface/seabed before arriving at a receiver, creating secondary paths that are sometimes many times longer than the first arrival path, generating symbol spreading over orders of seconds depending on the ranges and depths involved. Extensive Forward Error Correction coding is used on such channels to minimise packet losses.

$$A_{\mathrm{RF}}(d,f) \approx \left(\frac{4\pi df}{c}\right)^2 \text{ where } c \approx 3 \times 10^8 ms^{-1} \tag{3.4}$$

Thus, the multi-path channel transfer function can be described by

$$H(d,f) = \sum_{p=0}^{P-1} h(p) = \sum_{p=0}^{P-1} \Gamma_p / \sqrt{A(d_p,f)} e^{-j2\pi f \tau_p} \tag{3.5}$$
$$\text{where } \tau_p = d_p/c, c \approx 1500 ms^{-1}$$

where $d = d_0$ is the minimal path length between the transmitter and receiver, $d_p, p = \{1, \dots P-1\}$ are the secondary path lengths, Γ_p models additional losses incurred on each path such as reflection losses at the surface interface, and $\tau_p = d_p/c$ is the delay time ($c \approx 1500 ms^{-1}$ is the nominal speed of sound underwater).

Comparing $A_{aco}(d,f)$ with the RF Free-Space Path Loss model $A_{\mathrm{RF}}(d,f) \approx \left(\frac{4\pi df}{c}\right)^2$, the impact of range on signal power is exponential underwater, rather than quadratic in RF space ($A_{\mathrm{aco}} \propto f^{2d}$ vs $A_{\mathrm{RF}} \propto (df)^2$). While both frequency dependant factors are quadratic, approximating the factors in (3.3), $f \propto A_{\mathrm{aco}}$ is at least 4 orders of magnitude higher than $f \propto A_{\mathrm{RF}}$

3.0.6 Trust in Marine Networks

With demand for smaller, more decentralised marine survey and monitoring systems, and a drive towards lower per-unit cost, TMFs are going to be increasingly applied to the marine space, as the benefits they present are significant. Beyond the constraints of the communications environment, knock on pressures are applying in battery capacity, on-board processing, and locomotion. These pressures simultaneously present opportunities and incentives for malicious or selfish actors to appear to cooperate while not reciprocating, in order to conserve power for instance. These multiple aspects of potential incentives, trust, and fairness do not directly fall under the scope of single metric trusts discussed above, and this context indicates that a multi-metric approach may be more appropriate.

Chapter 4

Trust in Autonomous Systems of Systems for Maritime Defence Applications

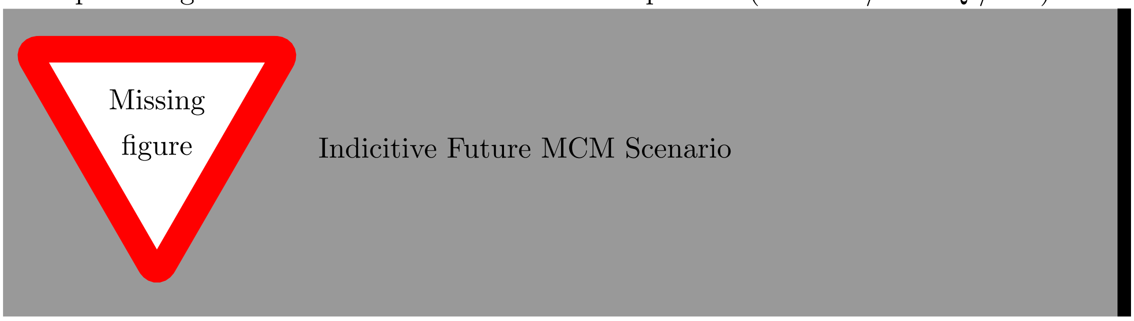
The aim of the chapter is to explore where trust is likely to impact on an indicative system (of systems) that contains autonomous elements. To assist with scoping this, an indicative scenario is selected from the Maritime domain. This scenario centres on autonomous Mine Counter Measures and/or Hydrography, Capability (MCM/MHC) operations, incorporating Human Factors, Command and Control (C2) concerns, and Vehicle to vehicle (V2V) distributed communication, from the perspective of trusted and semi-trusted operation.

With demand for smaller, more decentralised marine survey and monitoring systems, and a drive towards lower per-unit cost, TMFs are going to be increasingly applied to the marine space, as the benefits they present are significant. Beyond the constraints of the communications environment, knock on pressures are applying in battery capacity, on-board processing, and locomotion. These pressures simultaneously present opportunities and incentives for malicious or selfish actors to appear to cooperate while not reciprocating, in order to conserve power for instance. These multiple aspects of potential incentives, trust, and fairness do not directly fall under the scope of single metric trusts discussed above, and this context indicates that a multi-metric approach may be more appropriate.

However, the implications of trust in autonomy beyond securing communications and data are an area in need of further research (BAE Systems, 2013. Maritime Autonomy Final Report - Combined Response,) Of particular concern is the verification of autonomous behaviours. Technology Readiness Level deficiencies were identified in the Maritime Capability Contribution of Unmanned Systems (MCCUS) Osprey Phase 1 report(Clark, H. et al., 2012. Maritime Capability Contribution of Unmanned Systems,), with a particular focus on failsafe behaviour. The addition of increased on-board autonomy in MUxS, properly understood and verified, would greatly improve this future capability, similar to recent developments in the UAS arena[?]. Under the Osprey concept of operations, there is an opportunity for increased decentralisation and in-field collaboration(Walton, R., 2012. Maritime Autonomy PDR Pack.), however, difficulties in Trust between human operators and autonomous systems have already been clearly identified[?],and this has been demonstrated by the recent decision by the German government to renege on its 500M investment in the Euro Hawk programme, due to concerns about civil certification of the onboard autonomy[?] In order for these new distributed structures to be relied upon to provide operational performance, reliability and to maintain in-field situational awareness, vulnerabilities to disruption, interruption, and subversion need to be understood and minimised.

In order to contextualise the discussions on trust in mixed and hybrid networks, an exemplar scenario is considered. That scenario builds on existing Maritime Autonomy Framework (MAF) investigations(Mollet, J. et al., 2012. Osprey Task 37 Activity 8 - Unmanned Systems Operations: Technical Assurance Work Package - Security Issues and Mitigations - Final Report,)

While the initial assessment does not cover the MHPC PT CONUSE recommendations, it provides a starting point for future trust research in UxV operations. In order to constrain the scope of this project, a single operational scenario will be analysed within documented MCHP CONUSE(Rudge, A., Chapman, K. & Goddard, N., 2012. Information Management for MHPC: Research Strategy,), of Route/Area Survey within both peacetime and wartime contexts, with a Beyond Line of Sight (BLOS) operator. This scenario will be a minimal MCM operation in a littoral area. In field assets will consist of: Two squads consisting of Three UUVs, (tacitly modelled on the in-service REMUS 100 UUV), and a USV providing acoustic-RF relay capabilities per-squad an UAV providing BLOS Comms A remote human operator (MCMV / PJHQ / etc)



The differential between the peacetime and wartime contexts will be an attempted capture of a UUV by a manned surface-based FIS asset. Clearly, this paper has a limited scope and does not attempt to cover every aspect of a trustworthy system.

4.1 Trust Perspectives

In Human trust relationships it can be seen that there can be several perspectives of Trust for example organizational, sociological, interpersonal, psychological and neurological[6]. For the purposes of this work we can define two perspectives: Design and Operational. These are summarised as follows:

- Design Trust. When an autonomous system is under development a level of Trust is established in it through the manner in which it has been designed and tested. This is the same as conventional systems. The difference with systems that have high-levels of autonomy is that they are designed to behave adaptively to dynamic environments that are difficult to fully predict prior to operational deployment. For example, in a navigation system it is difficult to predict the dynamic environment it will need to adapt to. So Trust needs to be developed that the design and test of such systems are sufficient to predict that operational solutions will be, if not optimal, at least satisfactory.

- Operational Trust. Effectively, there are two aspects to this: trust that a system is operating as expected (which is inevitably tied in with, but distinct from) Design Trust; and trust that the interfaces between the operator and the system are as expected. This latter aspect covers issues such as physical links and interpretation of data.

Examples of roles that interact with a system from both of these trust perspectives are provided in Table 1 and Table 2 below.

Table 4.1: Examples of Roles that require a Design Perspective of Trust in Autonomous Systems.

	Role		
	Designer	**Acquirer**	**Disposer**
Definition	Responsible for developing the system	Responsible for acquisition of the system	Responsible for the disposal of a system.
Level	Organisation	Organisation	Organisation
Perspective	The designer of an Autonomous System develops trust through the application of known and trusted tools to well understood problems (e.g. a well-defined requirement set) using competent and trusted staff. The trust perspective therefore could be regarded as the **Design perspective**.	The Acquirer of a System develops trust through prior experience of the vendor and similar products. For any given product this is supplemented by the examination of engineering evidence provided by the Designer Organisation. Although there will be several trust aspects to the role, for the purposes of this paper this role can be seen as having a **Design perspective** since the Acquisition process needs to develop trust that the systems it is buying will be designed to be trustworthy in operation.	System disposal does not necessarily indicate destruction. Where assets are passed to 3rd parties (e.g. though sale) the disposer must be confident that the autonomous behaviour can be reduced (where necessary) to a known and acceptable level. This perspective is therefore part of the **Design perspective** since there will be trust that (possibly advanced) behaviours can be prevented from being passed unwittingly to second user organisations; particularly since they may use the systems in a different context.

Table 4.2: Examples of Roles that require a Operational Perspective of Trust in Autonomous Systems.

	Role		
	Commander	**Operator**	**User**
Definition	Responsible for the system tactical activity (e.g. mission / activity setting)	Responsible for the on-going control of the system when deployed on a particular mission / activity	An end user of the capabilities provided by the system.
Level	Person	Person	Person/System/Org.
Perspective	The Commander places trust in the acquisition process to provide reliable assets. However, their trust perspective is **operational**.	An operator develops initial trust in a system through training and experience of similar systems. When interacting with a deployed system, the ongoing trust is maintained through correct and understandable system behaviour. This can be regarded as **Operational Trust**	A user of a Systems capability may not have any knowledge of the System itself but will need to develop trust in ability to provide trustworthy services. Again, this may be regarded as a form of **Operational Trust**

4.1.1 Design Trust

Five aspects of Design Trust have been identified:

1. **Formal Specification of Dynamic Operation**: Autonomous Systems (AS) may be required to operate in complex, uncertain environments and as such their specification may need to reflect an ability to deal with unspecified circumstances. This includes engaging with dynamic systems of systems environments where an autonomous system may cooperate with a system not envisaged at design time. *How can systems that are required to demonstrate that they meet their requirement be specified flexibly enough to permit adaptive behaviours*?

2. **Security**: Any unmanned system has the potential to be used for illegitimate purposes by unscrupulous 3rd parties who could exploit security vulnerabilities to gain control of the system or sub-systems. Any system that has the potential to cause harm from such actions must have security designed in from the start to ensure that the system can be trusted to be resilient from cyber attack. Current accreditation

schemes rely on a security assessment of a known architecture and there are mutual accreditation recognition schemes that could be encoded in dynamic discovery handshake protocols. This would produce a secure network assured through the accreditation of its component systems. For example, the Multinational Security Accreditation Board (MSAB) deals with Combined Communications Electronics Board (CCEB) and NATO Accreditations to provide security assurance of internationally connected networks. Encoding such agreements into secure handshakes could enable dynamic accreditation of autonomous systems cooperating in a coalition environment. It is not known whether these have been demonstrated, so the question is: *Can autonomous systems be designed to understand the security situation when interfacing with known or unknown systems?*

3. **Verification and Validation of a Flexible Specification**: Following on from the description of a flexible specification, establish that the AS conforms and performs in accordance to the specification. This has direct implication for the trust in the resultant system. How can systems demonstrate that they will behave acceptably when the environment is unknown?

4. **Trust Modelling and Metrics**: This could be argued as part of the Verification and Validation of the system. However, models are increasingly being embedded into system design as a reference. Thus it is useful to consider this element separately. *How can trust be modelled sufficiently to span the space of most potential behaviours to help ensure that systems will be trusted when moved into operational environments? Can this be measured to allow comparison and minimum requirements set?*

5. **Certification**: The certification requirements placed on specific systems will vary depending on domain and national approaches to certification. However, the common element in the requirement for certification is that a certified system is deemed as sufficiently trustworthy for use within its context of certification. Additionally Certification also relies on the predictability of a system. Because the aim of autonomous systems is to deal effectively with uncertain environments, *can they (autonomous systems) be certified without being demonstrated in the environment within which they will adapt new behaviour?*

Clearly existing military and commercial standards can play a significant role in demonstrating the trustworthiness of any systems design. That is if a system has been designed to a Standard then it has known properties that have been accepted as good practice. However, these do not address the issue of the five areas listed above. The following sub section briefly outlines existing Standards for reference.

Current Unmanned System Interface Standardisation

There are three main organisations that are developing or have developed assurance standards for Unmanned Systems;

Table 4.3: Levels of Interoperability for STANAG 4586 Compliant UCS

LOI	
1	Indirect receipt/transmission of UAV related payload data
2	Direct receipt of Intelligence, Surveillance and Reconnaissance (ISR) data where direct covers reception of UAV payload data by the UCS when it has direct communication with the UAV
3	Control and monitoring of the UAV payload in addition to direct receipt of ISR/other data
4	Control and monitoring of the UAV, less launch and recovery
5	Launch and Recovery in addition to LOI 4

- NATO Standardization Office (NSO)
- Society of Automotive Engineers (SAE)
- American Society of Testing and Materials (ASTM)

NATO Standardization Office Faced with the growing adoption of similar but disparate UAV systems within NATO territories and coalition nations, STANAG 4586[**?**] , promulgated in 2005, defined a logistic and interoperability framework to provide commonality in the C2 architecture and implementations of UAV/Ground station communications.

This included a particularly interesting development in the form of "Vehicle Specific Module" (VSM) interoperability, whereby existing systems could be grandfathered into 4586 compliance by the addition of a VSM to operate as a protocol translator. This VSM could be mounted on the remote system, utilising a 4586 compliant Data Link Interface (DLI), or mounted on the UCS utilising a proprietary DLI to the remote system. 4586 described five Levels of Interoperability (LOI) for compliant UAV systems, shown in Table 3. This structure has been criticised as being short sighted and at odds with the reality of modern and proposed autonomous vehicle operations [**?**], specifically that in modern autonomous systems, there is no such thing as direct control or Operator-in-the-loop, especially in the case of BLOS systems, and that in increasingly autonomous systems, operation is done as Human Supervisory Control (HSC), or more commonly described as Operator-on-the-loop, whereby the operator interacts with the intermediate autonomous system and that autonomous system eventually performs that task on the hardware.

SAE Levels o
tonomy possi
from [**?**]

Further, 4586 predominantly deals with a 1-to-1 mapping between operators and assets, when this is quite against the current state of the art; greater focus is being made in collective and collaborative assignment and having a single operator managing a task force of assets in-field, and handing off vehicle management responsibilities to the individual assets.

Society of Automotive Engineers (SAE) The AS-4 steering group is responsible for the development and maintenance of the Joint Architecture for Unmanned System (JAUS) standards, which provide several service sets for Inter-System cooperation and interoperability, either in the form of a specified design language (JSIDL[1]) or as a direct framework implementation, such as the JAUS Mobility, Mission Spooling, Environment Sensing, or Manipulator Service Sets[2].

This provides a stack-like interoperability model akin to the OSI inter-networking standard, providing logical connections between common levels across devices regardless of how subordinate layers are implemented.

Importantly, JAUS service models are open-sourced under the BSD-license, and a development toolkit is available for anyone to develop JAUS-compatible communications and control protocols[**?**].

It is also important to note that JAUS is part funded, and heavily utilised by, US Army and Marine Robotic Systems Joint Project Office (RS-JPO), which manage the development, testing, and fielding of unmanned (ground) systems for those respective forces. This includes now legacy M160 mine clearance platform and the highly popular (both with forces and their in-field operators) iRobot Packbot inspection and EOD clearance family of robots.

American Society of Testing and Materials (ASTM) The ASTM F38 committee has developed a LoS, single-asset-single-operator stove-piped framework for Unmanned Air Systems that is too constrained in scope for applicability to a more heterogeneous operating environment[**?**]. However, the F41 Committee, focused on Unmanned Maritime Vehicle Systems (UMVS) has collectively developed a range of interoperable standards, covering Communications, Autonomy and Control, Sensor Data Formats, and Mission Payload Interfacing. Of particular interest is the Autonomy and Control standard [**?**], which highlighted a requirement on the vehicle system to be able to recognise an authorised client, be that a human operator or an additional collaborating vehicle. Further, the standard states that the responsibility of the safety and integrity of any payload remains with the vehicle. This standard was withdrawn in 2015 due to ASTM regulations requiring standards to be updated within 8 years of approval, and has no direct replacement within ASTM, but stands as a useful guiding perspective on autonomy standards within industry.

[1] JAUS Service Interface Definition Language

[2] SAE AS6009, AS 6062, AS 6060, and AS 6057 respectively

ASTM F41 UMVS Architecture

4.1.2 Operational Trust

This work is considering autonomous systems as entities of wider systems, we refer to these here as Autonomous Collaborative Systems. As described earlier, Operational Trust has two main aspects, trust in the system to behave as expected and trust in the interfaces between systems (human/machine and machine/machine). Of all of the interfaces in an Autonomous Collaborative System, the most problematic is that arguably that between the System of Autonomous Systems (SoAS) and the human operator / team of operators. Cummings identified the main challenges to Human Supervisory Control (HSC), summarised below:[**?**]

Information Overload

Operator efficiency exhibits an optimum at moderate levels of cognitive engagement, above which cognitive ability is overloaded and performance drops (Otherwise known as the Yerkes-Dodson Law). Additionally, in the case of under-engagement, operators can fall foul of boredom, and become desensitised to changing factors. *However, predicting this point of over-saturation is an open psychophysiological research problem.*

Adaptive Automation

Automation is well tailored to consistent levels of activity. This is quite simply not the case in the military domain, characterised by long periods of routine punctuated by high intensity, usually unpredictable, activity. At those interfaces between calm and storm, where SA and IA are imperative, temporary Information Overload is highly probable. Adaptive Automation enables autonomous systems to increase their level of automation (LOA) based on specific events in the task environment, changes in operator performance or task loading, or physiological methods. It is taken as given that for routine operations, and increased LOA reduces operator workload, and vice versa. However, this relationship is highly task dependent and can create severe problems in cases of LOA being greater, or indeed lesser, than is required. In the cases of overly-high LOA, operator skill is degraded, situational awareness is reduced as the operator is not as engaged, and the automated system may not be able to handle unexpected events, requiring the operator to take over, which, given the previous points, is a difficult prospect.

Alternatively, in sub-optimal LOA, Information Overload can result in the case of high intensity situations, but also the system can fall foul of overly-sensitive human cognitive biases, false positive pattern detection, boredom, and complacency in the case where less is going on. Therefore, as a corollary to Information Overload challenges, there is a need to define the interrelationship between levels of situational activity (or risk) and appropriate levels of automation. *Under what circumstances can AA be used to change the LOA of a system? Does the autonomous system or the human decide to change LOA? What LOAs are appropriate for what circumstances?*

Distributed Decision Making

In a modern, non-hierarchical, often distributed or cellular military management system (Network Centric Warfare doctrine for example), tools are increasingly being used to mitigate information asymmetry within C2. A simple example of this is shared watch-logs in the Naval space, providing temporal collaboration between watch-teams separated in time. The DoD Global Information Grid is another example of a spatial collaborative framework. Recent work has demonstrated the power of collaborative analysis and human-machine shared sensing technologies even with low levels of training on the part of the operators providing superior results and resource efficiencies than either humans or machines alone in survey and search-and-rescue scenarios (Ahmed et al. 2014). As these temporal and spatial collaboration tools increase in complexity and ability, decisions that previously required SA that was only available at higher echelons within the standard hierarchy are available to commanders on the ground, or even to individual team members, enabling the potential for informed decisions to be taken faster and more effectively, enabled by automated strategies to present relevant information to teams based on the operational context. However there are a range of operational, legal, psychological and technical challenges that need to be addressed before confidence in these distributed management structures can be established. Studies into SA sharing techniques (telepresent table-top environments, video conferencing, and interactive whiteboards) have generally yielded positive results, however investigations into interruptive-communications (such as instant messaging chat) have demonstrated a negative impact on operational efficiency. In short, the biggest problem with distributed decision making in the context of supervisory systems is that *there is no consensus on whether it is advantageous or not, and what magnitude of operational delta is introduced, if any.*

Complexity

Beyond simple Information Overload, increasing complexity of information presented to operators is having a negative effect on operational efficiency. In HSC, displays are designed to reduce complexity, introducing abstractions with an aim to presenting the minimum amount of information to the operator required to maintain an accurate and up-to-date mental model of the environmental and operational state. This has led to the development of many domain specific decision support interfaces, however, in academic research, there has been nothing but mixed results. One commonly raised negative is the general bias on the cool factor of interfaces. Immersive 3D visual, aural, or haptic interfaces that at first appraisal seem to provide more approachable information to the operator, and are indeed tacitly preferred by operators in use. However, there has not been any evidence to demonstrate performance improvement when using these tools, and in-fact, *improving the fidelity of the interfaces has led to operators overly-relying on these representations of the environment rather than remaining engaged in the environment.*

Cognitive Biases and Failing Heuristics

The increasingly connected battlefield has massively increased the tempo of operations, with increasing requirements on commanders and operators to make rapid decisions with imperfect information. However, Human decision making isnt always rational (especially under pressure), and operators use personally derived heuristics to make rational short-cuts. This is a double edged sword, where these heuristics can be employed to greatly reduce the normative cognitive load in a stressful situation, but also introduce destructive biases, where these shortcuts make assumptions that dont bear out in reality.

For example, in the context of decision support systems, Autonomy Bias has been observed as a complement to the already well known Confirmation Bias[3] and Assimilation Bias[4] , where operators that have been provided with a correct answer by a decision support system do not look (or see, depending on your perspective) for any contradictory information, and will unquestionably follow, increasing error rates significantly.

This behaviour isnt only the reserve of decision support systems, but also in the generic allocation of operator attention; scheduling heuristics are used to decide how much time tasks should be worked on, and time and again, humans are found to be far from optimal in this regard, especially in time-pressured scenarios where these heuristics are in even more demand. Even when operators are given optimal scheduling rules, these quickly fall apart, often due to primary task efficiency degradation after interruption. This highlights a critical interface in the adoption of complex autonomous systems that still demand Man in the loop functionality; if a system is required to have full-time concentrated supervision (e.g. flying a UCAV), but also event-based reactive decision making (e.g. alerts from non-critical subsystems), both tasks are negatively impacted. In an assessment of factors influencing trust in autonomous vehicles and medical diagnosis support systems, Carlson et al also identified that a major factor in an operator or users trust in a system was not only dependant on past performance and current accuracy but also on soft factors such as the branding and reputation of the manufacture /

[3]Confirmation Bias is the tendency for people to preferentially select from available information that information that supports pre-existing beliefs or hypotheses.

[4]Assimilation Bias is often thought of as a subset of Confirmation Bias, whereby it specifies that instead of seeking out information supporting of current views, any incoming data is interpreted as being supportive of a particular view without questioning that view, even if it appears contradictory.

designer.(Carlson et al. 2014) Further, autonomous decision support / detection / classification systems have an uncanny valley to overcome in terms of accuracy, in that there is a dangerous period when such systems are used but not perfect, but operators become complacent, causing an increased error rate, until such a time that those autonomous systems can match or exceed the detection rates of their human counterparts.

Summary of Human Factors impacting Operational Trust in Defence Contexts

When dealing with human supervision of autonomous or semi-autonomous systems, there is an inherent conflict between the expectations of the operator, the hopes of system architects. System Architects aim to provide more and more information to the operator to justify a systems operation, and Operators in reality need less and less information to be efficient when things are going well, and responsive in a dynamic environment. This places huge demands on Human Interface design and indeed on communications design to provide this timely, relevant, interactive connection between any autonomous system and the end operator(s). Recent work has presented the idea of taking user interface (UI) inspiration from the entertainment sector, in terms of UI best practises developed over two decades of Real-Time Strategy game development [**?**], and follow up work into automated mission debrief demonstrated that such operational support could improve causal situational awareness of an operator when compared to a human-baseline [**?**]. In terms of the human factors challenges raised by Cummings, they are often contradictory in their direction, particularly when contrasting between Adaptive Automation and Cognitive Biases challenges. This is a key part of the soft-trust theory, where the operators and commanders need to be able to implicitly and explicitly trust the operation of a remote system with limited feed-back bandwidth, high latency, or long-term operation such that direct remote operation is infeasible or undesirable. To be able to trust that systems ability to continue on a course, survey an area, notify on detection of an anomaly, etc. is going to be the corner stone of any autonomous systems justification in the future.

4.2 Trust and Reputation in Autonomous Collaborative Systems

In addition to the two perspectives of trust identified thus far, and for the purposes of this investigation, it is necessary to define and classify Operational Trust into two distinct but related sections, which we define as being

- Hard Trust or technical trust, being the quantative measurement and communication of the expectation of an actor performing a certain task, based on historic performance and through consensus building within a networked system. Can be thought of as a de-risking strategy to measure the ability of a system to perform a task unsupervised.

- Soft Trust or common trust, being the qualitative assessment of the ability of an actor to perform a task or operation consistently and reliably based on social or experiential factors. This is the natural form of trust and is the main motivational driver for the human-factors trust discussion. Can be rephrased as the level of confidence in an actor to perform a task unsupervised.

It is already clear that these two definitions are extremely close in their construction, but represent fundamentally different approaches to trust, one coming from a sociological perspective of person-to-person and person-to-group relationships from day to day life, and the other coming from a statistical appraisal of an activity by a system. The difficulty with human supervisory controlled autonomous systems is that there is a need for both a hard and soft trust perspective, and that this interface can often create fundamental misunderstandings.

4.3 Levels of Trust

Trust relationships operate as part of a system architecture, and can quite often get confused. As such, we constrain the focal domain as per [9] into six constructs. Sun[**?**] suggests that within these there are two overarching forms of trust:

- Behavioural: That one entity voluntarily depends on another entity in a specific situation
- Intentional: That one entity would be willing to depend on another entity

These concepts closely mirror the authors definitions of Hard and Soft trust respectively, one (Behavioural) being an invested dependency given certain parameters being satisfied, mirroring Hard Trust, and the other (Intentional) being the capacity for belief in another entity, analogous to Soft Trust. It is suggested that these overarching forms are supported by and indeed are drawn from four major constructs within social and networked environments:

- Trusting Belief: the subjective belief within a system that the other trusted components are willing and able to act in each-others best interests
- Dispositional Trust: a general expectation of trustworthiness over time
- Situational Decision Trust: in-situ risk assessment where the benefits of trust outweigh the negative outcomes of trust
- System Trust: the assurance that formal impersonal or procedural structures are in place to ensure successful operation.

While Sun argues that only System Trust and Behavioural Trust are relevant to trusted networking applications. However, it is arguable that in any network where the operation of that network is not the only concern, or where that network has to interact with any operator, then all of these factors come into play. Both System and Behavioural trust rely on what Sun calls a Belief Formation Process, or a trust assessment, while the other trust constructs deal with the interactions between trust and decision making against an internal assessment of network trustworthiness.

Chapter 5

Strategies for Multi-Domain Trust Assessment

Chapter 6

Modelling and Analysis of Collaborative Node Kinematic Behaviuors in Underwater Acoustic MANETs

6.0.1 Establishing Scale Factors in Communications Rate

In this section we characterise the simulated communications environment, establishing an optimal packet emission rate for comparison against [5].

In order to establish the point at which the network becomes saturated due, a range of packet emission rates were explored between 0.01 packets per second (pps), equivalent to 96 bps, up to 0.07 pps (672 bps)

From Figs. 6.1 and 6.2, it is clear that the threshold curve, expressed as the *Successfully Received Packets* line, exhibits a saturation point between 0.025 and 0.03 pps. Particularly in Fig. 6.2, the precipitous drop in packet delivery probability beyond 0.025 pps, indicating that this is a strong candidate value for an upper-limit to the safe operating zone in terms of packet emission in the small static case.

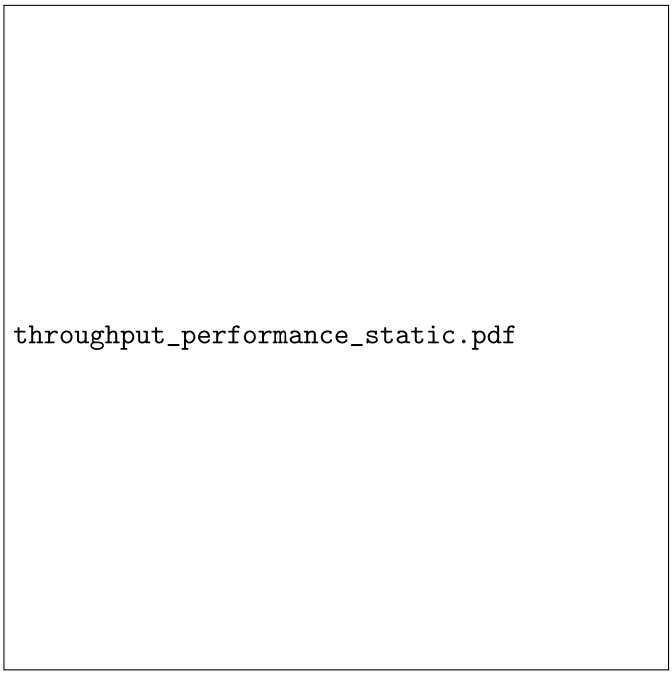


FIGURE 6.1: Varying packet emission rate demonstrates maximal throughput at 0.025 packets per second, equivalent to ≈240 bps

FIGURE 6.2: Varying packet emission rate demonstrates a saturation point at 0.025 packets per second

6.0.2 Establishing Scale Factors in Physical Distribution

In this section we characterise the effect of node-separation scaling on communications operation for comparison against [5]. This is particularly important considering the significant scale factor differences between not only the speed of propagation in the medium, but simply the range of operation. From Table **??**, the operating transmission range of acoustic is ≈ 6 times further than 802.11, indicating that a suitable operating environment will have an area $\approx \sqrt{6}$ times the area of the 802.11 case. Therefore, a reasonable experimental range would have an upper bound of performance around this scaling factor, where nodes are approximately $400m$ apart.

A reasonable range around this is to scale from $100m$ apart on average to $800m$.

Varying average node separation shows that while direct throughput isn't significantly affected until, collision rates are Fig. 6.3. This collision rate is well within the tolerances of the MAC layer, as shown in Fig. 6.4, where even with a rising collision rate, packets are being reliably received.

`throughput_performance_range.pdf`

FIGURE 6.3: Comparison of Medium Acquisition Collisions, Throughput, and Enqueued packets against varying application packet emission rates.

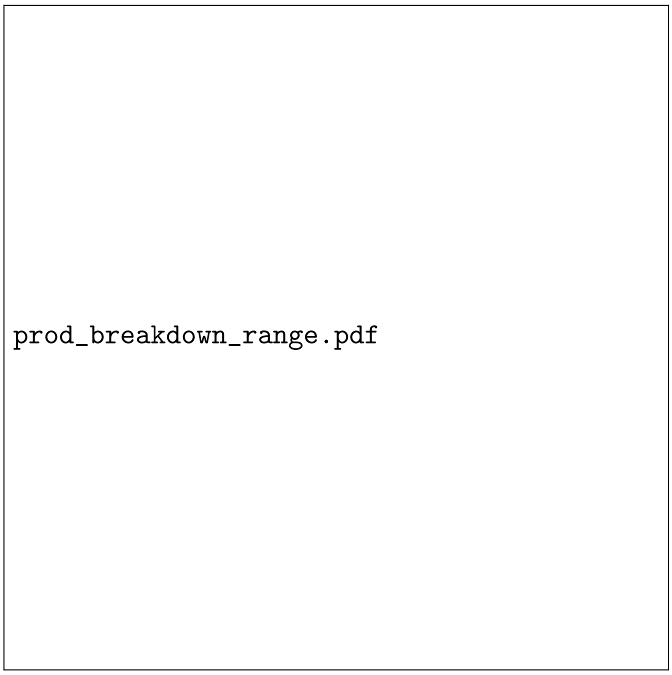


FIGURE 6.4: Probability of Timely Reception across a range of node scaling.

However, when end-to-end delay is investigated, it's clear from Fig. 6.5 that the network is becoming severely impaired approaching the $600m$ mark, with delays rising to more than 25 minutes above $700m$. This is also demonstrated by the increasing RTS/Data ratio shown in Fig. 6.6.

According to Xu [17], the RTS/CTS handshake cannot function well as interference protection at node separations beyond 0.56 times the transmission range. This is also demonstrated in Fig. 6.6, where above $1500m \times 0.56 = 840m$, This is due to reduced channel availability due to collisions, which are then due to a much longer potential contention period between nodes.

delay_range.pdf

FIGURE 6.5: End to End Delay under varying node-separations

rts_range.pdf

FIGURE 6.6: RTS/Data ratio for varying node-separations

TABLE 6.1: Tabular view of data from Figs 6.4, 6.5, and 6.6

Separation(m)	Delay(s)	Probability of Arrival	RTS/Data Ratio	Ideal Delivery Time(s)
100	60.32	0.99	1.80	1.03
200	419.95	0.97	2.02	1.10
300	1205.66	0.89	2.41	1.17
400	1288.20	0.91	2.26	1.25
500	1868.20	0.87	2.41	1.32
600	2191.07	0.85	2.42	1.39

6.0.3 Metric Weighting

./Figures/trust_bella_static_selfish.pdf

(A) All Nodes Static

trust_bella_single_mobile_selfish.pdf

(B) n_1 Randomly Walking

trust_bella_allbut1_mobile_selfish.pdf

(C) All Nodes but n_1 Randomly Walking

/Figures/trust_bella_all_mobile_selfish.pdf

(D) All Nodes Randomly Walking

FIGURE 6.7: MTFM Trust assessments for varying mobility options in the selfish case

beta_trust_bella_static_joint.pdf

(A) All Nodes Static

beta_trust_bella_single_mobile_joint.pdf

Chapter 7

Comparative Analysis of Multi-Domain Trust Assessment in Collaborative Marine MANETs

Bibliography

[1] Andrew Bolster and Alan Marshall, *Single and Multi-Metric Trust Management Frameworks for use in Underwater Autonomous Networks*, TrustCom2015.

[2] Sonja Buchegger and Jean-Yves Le Boudec, *Performance analysis of the CONFIDANT protocol*, Proc. 3rd ACM Int. Symp. Mob. ad hoc Netw. Comput. - MobiHoc '02, ACM Press, 2002, p. 226.

[3] Andrea Caiti, *Cooperative distributed behaviours of an AUV network for asset protection with communication constraints*, Ocean. 2011 IEEE-Spain (2011).

[4] Jin-hee Cho, Ananthram Swami, and Ing-ray Chen, *A survey on trust management for mobile ad hoc networks*, Commun. Surv. & Tutorials **13** (2011), no. 4, 562–583.

[5] Ji Guo, Alan Marshall, and Bosheng Zhou, *A new trust management framework for detecting malicious and selfish behaviour for mobile ad hoc networks*, Proc. 10th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. Trust. 2011, 8th IEEE Int. Conf. Embed. Softw. Syst. ICESS 2011, 6th Int. Conf. FCST 2011 (2011), 142–149.

[6] John D Lee and Katrina A See, *Trust in automation: designing for appropriate reliance.*, Hum. Factors **46** (2004), no. 1, 50–80.

[7] Huaizhi Li and Mukesh Singhal, *Trust Management in Distributed Systems*, Computer (Long. Beach. Calif). **40** (2007), no. 2, 45–53.

[8] Jie Li, Ruidong Li, Jien Kato, Jie Li, Peng Liu, and Hsiao-Hwa Chen, *Future Trust Management Framework for Mobile Ad Hoc Networks*, IEEE Commun. Mag. **46** (2007), no. 4, 108–114.

[9] K J R Liu, *Information theoretic framework of trust modeling and evaluation for ad hoc networks*, IEEE J. Sel. Areas Commun. **24** (2006), no. 2, 305–317.

[10] Sifeng Liu and Yi Lin, *Grey System Theory and Application*, no. 1, Springer-Verlag Berlin Heidelberg, 2011.

[11] Junhai Luo, Xue Liu, Yi Zhang, Danxia Ye, and Zhong Xu, *Fuzzy trust recommendation based on collaborative filtering for mobile ad-hoc networks*, 2008 33rd IEEE Conf. Local Comput. Networks (2008), 305–311.

[12] M E G Moe, B E Helvik, and S J Knapskog, *TSR: Trust-based secure MANET routing using HMMs*, ...symposium QoS Secur. ... (2008), 83–90.

[13] David K W Ng, *Grey System and Grey Relational Model*, SIGICE Bull. **20** (1994), no. 2, 2–9.

[14] Jim Partan, Jim Kurose, and Brian Neil Levine, *A survey of practical issues in underwater networks*, Proc. 1st ACM Int. Work. Underw. networks WUWNet 06 **11** (2006), no. 4, 17.

[15] Milica Stojanovic, *On the relationship between capacity and distance in an underwater acoustic communication channel*, 2007, p. 34.

[16] Y Wang, V Cahill, E Gray, C Harris, and L Liao, *Bayesian network based trust management*, Auton. Trust. ... (2006), no. 60373057, 246–257.

[17] Kaixin Xu, Mario Gerla, Sang Bae, and Hoc Networks, *Effectiveness of RTS / CTS Handshake in IEEE*, ..., 2002. Globecom'02. Ieee **56** (2002), 1–14.

[18] Fengchao Zuo, *Determining Method for Grey Relational Distinguished Coefficient*, SIGICE Bull. **20** (1995), no. 3, 22–28.