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

December 2015

Contents

Illustrations

List of Figures

List of Tables

Glossary

ACS Autonomous Collaborative System. vi, 17

AUV Autonomous Underwater Vehicle. vi, 28, 34, 38

BLOS Beyond Line of Sight. vi, 16

CMRE Centre for Maritime Research and Experimentation. vi, 40

DSTL Defence Science and Technology Laboratory. vi, 40

EOD Explosive Ordnance Disposal. vi, 17

HRI Human Robot Interaction. vi, 10

HSC Human Supervisory Control. vi, 16, 18

ISR Intelligence, Surveillance and Reconnaissance. vi, 16

JAUS Joint Architecture for Unmanned Systems. vi, 17

LOA Level of Automation. vi, 11, 12, 18, 19

LOI Level of Interoperabilty. vi, 16

LOS Line of Sight. vi, 17

MANET Mobile Ad-hoc Network. i–iii, v, vi, 1–28, 34, 37–39, 46, 49, 53–55, 58

MCM Mine-Counter Measure. vi

MHC Maritime Hydrography Capability. vi

MTFM Multi-parameter Trust Framework for MANETs. vi, 24, 26, 27, 34, 36, 37, 39

OTMF Objective Trust Management Framework. vi, 23–25, 27, 38, 39

PKI Public Key Infrastructure. vi, 2, 22

PLR Packet Loss Rate. vi, 23, 39

SoS System of Systems. vi, 9

TMF Trust Management Framework. vi, 3, 4, 38, 39

TTP Trusted Third Party. vi, 2

UAN Underwater Acoustic Network. ii, vi, 34, 38

UAV Unmanned Aerial Vehicle. vi, 9, 16

UMVS Unmanned Maritime Vehicle System. vi, 17

V2V Vehicle to Vehicle. vi, 9

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, even though I swore I'd never do it, this work is dedicated to R, who knows why.

Alan-hu Akbar

Chapter 1

Introduction

1.1 Mobile Ad-hoc Networks (MANETs)

With the explosive growth in the use of mobile telephony and the increasing miniaturisation and efficiency gains of portable communications devices, the classical paradigm of a broadcast/receiver or server/client has given way to an increasing use of decentralised, ad-hoc networks that not only accommodate but take advantage of network mobility.

Wether these networks are decentralised cellular / RF / 802.11 WiFi networks for use in disaster relief areas [1] or biologically inspired wireless sensor networks for low-energy, low-maintenance environmental monitoring [2],MANET theory developed over the past 30 years has gone from it's first formal definition, emerging from DARPA's Packet Radio Network research[3], to being an integral part of modern practical communications.

Inappropriat Citation

Minimally, a MANET consists of of a collection of mobile physical entities (nodes) with some form of communications, processing/data collection, and power systems. Similarly in terms of communications capability, while in many cases MANET nodes incorporate bi-directional transceivers to send and receive data, this bi-directionality is also not a limiting factor on inclusion within the MANET field, particularly in the area of Wireless Sensor Networks [4]. Nor is the capability or mix of communications technologies used; omnidirectional, static, or steerable communications antennae, utilising a range of technologies such as WiFi, Bluetooth, GSM, UMTS, Optical or Acoustics. A core characteristic of the design of MANETs is the inclusion and integration of heterogeneous node collections, i.e where different nodes or groups of nodes in a network have difference capabilities, whether this be in terms of propulsion, sensor apparatus, communications capability, etc.

These networks may be totally independent with no external connections, include independent per-node communications backhauls (e.g. Cellular Modems in mobile phones as part of a Bluetooth Personal Area Network(PAN)), or include static nodes that provide infrastructure based backhaul. However, this multiplicity of variations and options presents several challenges to users and operators; the physical topology of MANETs can vary wildly over short periods of time. A particular challenge to MANET operation is that given any node may operate as a routing / gateway node, if/when that node

Dynamic Topologies	Nodes are free to move arbitrarily; thus, the typically multihop network topology may change randomly and rapidly at unpredictable times, and may consist of both bidirectional and unidirectional links.
Bandwidth Constrained, Varied Capacity	Wireless links will continue to have significantly lower capacity than their hardwired counterparts. In addition, the realized throughput of wireless communications, after accounting for the effects of multiple access, fading, noise, and interference conditions, etc., is often much less than a radio's maximum transmission rate. One effect of the relatively low to moderate link capacities is that congestion is typically the norm rather than the exception, i.e. aggregate application demand will likely approach or exceed network capacity frequently.
Energy Constrained Operation	Some or all of the nodes in a MANET may rely on batteries or other exhaustible means for their energy. For these nodes, the most important system design criteria for optimization may be energy conservation.
Limited physical security	Mobile wireless networks are generally more prone to physical security threats than are fixed cable nets. The increased possibility of eavesdropping, spoofing, and denial-of-service attacks should be carefully considered. Existing link security techniques are often applied within wireless networks to reduce security threats. As a benefit, the decentralized nature of network control in MANETs provides additional robustness against the single points of failure of more centralized approaches.

TABLE 1.1: Summary of Characteristics of MANETs[5]

moves to a different region, network segments that had previously used that node as a path must renegotiate / reestablish their routes. These situations, if not appropriatly managed, lead to opportunities for subversion and selfishness.

The characteristics of MANETs as defined by Corson et al. are paraphrased in Table 1.1.

1.2 MANETs in Harsh Environments

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.

The distributed and dynamic nature of MANETs mean that it is difficult to maintain a trusted Trusted Third Party (TTP) or evidence based trust system such as Certificate Authorities or using Public Key Infrastructure (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.

possibly worthwhile doing more background on the operation of these

As such, Trust Management Frameworks (TMFs) 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). 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 [6], and maintaining throughput in the presence of malicious actors [7]

1.3 Systems Approach to Trust and Trust Engineering

1.4 Trust operation against Capable Attackers

1.5 Conclusion

Chapter 2

Background on Trust and its Applications to MANETs

In this chapter we explore the current literature and research around the concepts, theory, and application around Trust and Trust Management, specifically leaning towards the applications of Trust within Autonomous MANETs.

In the first section, the generic operations and background to Autonomy and "Trusted Operation" from a user/operators perspective is investigated. In the second section, the abstract quantity of "Trust" is explored. In the third section, current use and applications of Trusted operation of MANETs is explored, including current TMFs.

2.1 Trust Definitions and Perspectives

For a term that is so common in every-day speech, Trust is a challenging discussion area, particularly given the wealth of proposed definitions (Table 2.1). Beyond these dry, vague, and often "fuzzy" definitions, there is a significant ontological conflict between the subjective and objective perspectives of trust; is "trust" an attribute of the actor performing a given action, or of the observer of such an action? Or indeed is trust itself an action upon a relationship between actors? Is it qualitative or quantitative? These questions have challenged philosophers, psychologists and social scientists for decades.

More of these in the bookmarks list

In human trust relationships it is recognized that there can be several domains of Trust for example organizational, sociological, interpersonal, psychological and neurological [8].

These domains of trust are, from a human perspective, quite natural and are formed during the earliest stages of linguistic integration. This leads to recognisable deviations in the experiential concept of "trust" across cultures with differing linguistic histories. This has led to a wealth of work in the social sciences (as well as management schools across the world) in to how to develop, understand, and repair trust across cultural boundaries.[9]

Get more citations for this paragraph, need

As such it is important to explore the following areas of Trust definition before approaching the application of Trust towards Autonomous Systems and finally to MANETs:

Table 2.1: Definitions of Trust

Definition	Source
Assured reliance on the character, ability, strength, or truth of someone or something.	Merriam-Webster
Firm belief in the reliability, truth, or ability of someone or something	OED
The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a articular action important to the trustor, irrespective of the ability to monitor or control that other party	[10]
An expectancy held by and individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon	[11]

Table 2.2: Factors of Trust[10]

Factor	Definition
Ability	Collection of skills, competencies, capabilities and characteristics that enable a party to have influence or action within some specific domain
Benevolence	The extent to which a trustee is believed to want to do good to or by the trustor beyond a selfish profit motive
Integrity	Acceptance or adherence to a common set of principals of operation that the trustor finds acceptable

1. Definitions of Trust
2. Modelling and Analysis of Trust Relationships
3.

2.1.1 Modelling of Trust Relationships

Mayer et al [10] proposed a model of trust that encapsulates generalised factors of perceived trustworthiness in interpersonal relationships (Table 2.2), accommodating a subjective trustworthiness and risk-taking potentiality on the part of the trustor. This formulation of trust allowed a wider discussion of the characteristics of trust relationships, both between individuals and within networks or communities.

Lee and See [8] extended and synthesised Mayer et al's approach to personal and interpersonal trust towards a generalised concept of trust for human and autonomic/autonomous systems with the following alternative contextual definitions (including their approximate mappings to Mayer et al's approach

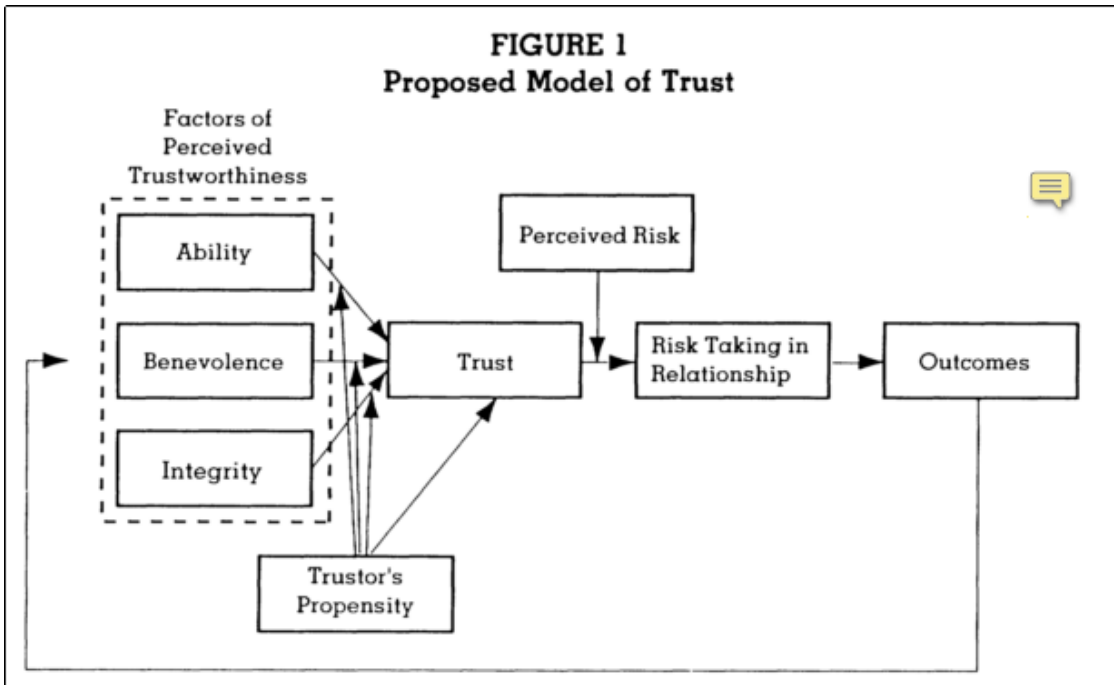


FIGURE 2.1: Model of Trust [10]

TABLE 2.3: Factors of Trust for Autonomous Systems[8]

Factor	Definition	Mayer Term
Performance	'The current and historical operation of the automation, including characteristics such as reliability, predictability, and ability	Ability
Process	The degree to which the automation's algorithms are appropriate for the situation and able to achieve the operators goals.	Integrity
Purpose	The degree to which the automation is being used within the realm of the designers intent	Benevolence

2.1.2 Characteristics of Trust Relationships

There are five commonly considered characteristics or attributes of Trust relationships in general, but not all relationships exhibit them and they are not assumed to be a complete specification of Trust:

- *Multi-Party* - One-to-one; one-to-many; many-to-one; many-to-many. Trust is not an absolute characteristic of a lone individual. Trust may include multi-agent abstractions (one-to-many), such as a preferential trust/distrust towards a group exhibiting a particular attribute, e.g. members of the armed forces / police services. Likewise, there can be trustor/trustee attributes that can generalise relationships between collectives (many-to-many), e.g. Jets and Sharks

- *Transitive* - Trust assessments can be shared (i.e. recommendations), where this second order trust assessment incorporates both the observed trustworthiness of the trustee, as well as the trustworthiness of the intermediate trustor. In some models this is further extended to include out-of-network intermediate trustors that have some other defined authority, e.g. PKI Certificate Authority

- *Evidential* - Trust must be based on some form of evidence-based observation or assessment, such as historical success rates of performing a certain action, or second-hand observations of trust from a third party.

- *Directional Asymmetry* - The majority of relationships are bi-directional but are asymmetric, i.e. between two entities who "trust" each other, there are two independent trust relationships that may have very different "values" or extents.

- *Contextual* - Trust can be variable and loosely coupled between contexts with respect to the action being assessed or the environment within which the trustee is operating, e.g. Doctors are trusted to perform medical procedures but that trust may not improve their success at correctly wiring an electrical plug. However there are plenty of counter-examples to this, as from [10], two of the three listed factors of trust are "Benevolence" and "Integrity" and are unrelated to the ability of a trustee to perform a particular action, so it is reasonable to make an initial assumption that if a trustee is being benevolent in one activity or context, that that benevolence *should* extend to other contexts.

2.1.3 Topologies of Multi-Party Trust Networks

Beyond the attributes or characteristics of an individual trust relationship, within any multi party sparsely connected network or community, topological context is useful in both establishing trust and in disseminating observations for collaborative assessment.

Within sparsly connected networks, there are three primary types of relationship, minimally demonstrated in Fig. 2.2;

- *Direct* - Whereby two nodes have a zero-hop communications link between them (A, B, C in the given figure)

- *Indirect* - Where two nodes have a $n > 1$ hop communications link (E, D from A or Cs perspective in the given figure)

- *Recommendation* - Where three nodes are fully connected so as to enable the exchange of direct opinions and form composite opinions based on the target and reporter (i.e. A may have both its own Direct assessment of C, as well as it's knowledge of Bs Direct assessment of C)

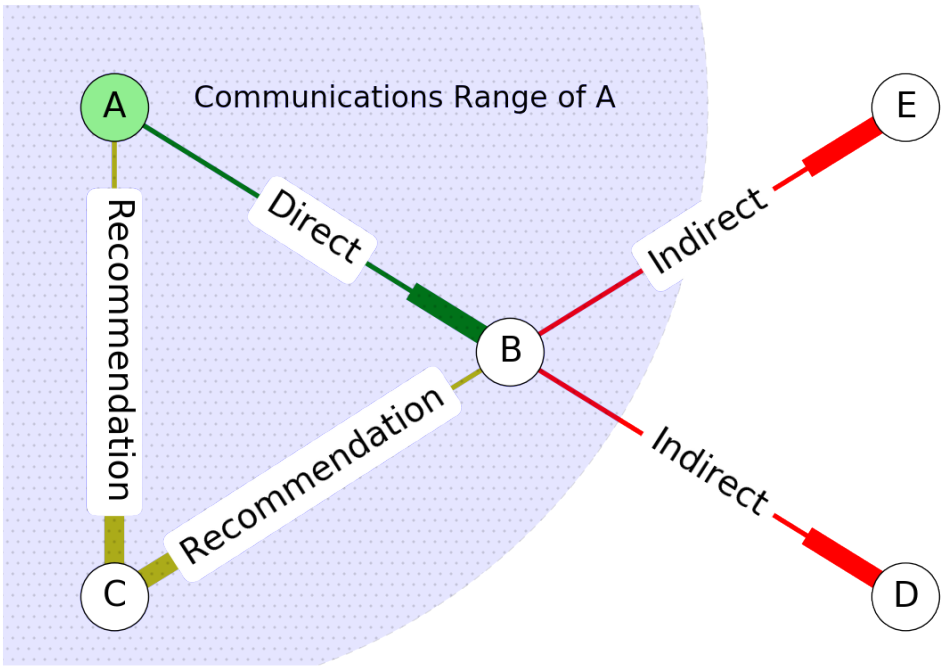


FIGURE 2.2: Trust Topologies, Direct, Indirect, Recommender, etc. from the perspective of Node A

2.1.4 Levels of Trust

This section may be superfluous

Trust relationships operate as part of a system architecture, and can quite often get confused. Sun[12] 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 previous definitions of Hard and Soft trust respectively. Behavioural Trust can be considered an invested dependency given certain parameters being satisfied, mirroring Hard Trust, and Intentional Trust can be considered as 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.

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.

2.2 Trusted Development and Operation of Autonomous Systems

2.2.1 Introduction

The aim of the section is to explore where trust is likely to impact System of Systems (SoS) that contain autonomous elements incorporating Human Factors, Command and Control considerations, and Vehicle to Vehicle (V2V) distributed communication, from the perspective of trusted and semi-trusted operation.

This is a subject area with a great deal of complexity and touches on a wide range of research areas, so it will be discussed in four parts;

1. Autonomy and Levels of Autonomy
2. Trust as a Design Consideration for the Development of Autonomous Systems
3. Operator / Organisational Factors relating to the Operation of Highly Autonomous Systems

2.2.2 Autonomy and Levels of Autonomy

Autonomy, like trust, is a nebulous, ill defined term across research, defense and commercial circles. In the most general case, Autonomy is explained as the allocation of functionality between a system and an operator tasked with performing a given task; where a system is more"autonomous", more of the sensing, planning, decision and action operations are performed by the system. (See Table 2.4 for a review of current definitions of autonomy and autonomous systems)

While Autonomy is largely taken to be a robotics term based in the case of one human operator and one robotic entity, recent advances in the development of more generalised cyber-physical systems has expanded this definition; from over-the-horizon human operation of Unmanned Aerial Vehicles (UAVs) to global networks of collaborating machines such as Google and beyond.

Possibly need to include discussion of the nature and levels of autonomy up here, but is a fairly common concept

TABLE 2.4: Definitions of Autonomy

Definition	Source
...should be able to carry out its actions and to refine or modify the task and its own behaviour according to the current goal and execution context of its task	Alami et al. [13]
Autonomy refers to systems capable of operating in the real-world environment without any for of external control for extended periods of time	Bekey [14]
...a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect with it senses in the future. ...Exercises control over its own actions	Franklin and Graesser [15]
An unmanned systems own ability of sensing, perceiving, analyzing, communicating, planning, devision-making, and acting, to achieve goals as assigned by its human operator(s) through designed HRI. ...The condition or quality of being self-governing	Huang [16]
...that the robot can operate self-contained, under all reasonable conditions without requiring recourse to the human operator. Autonomy means that a robot can adapt to change in its environment ...or itself ...and continue to reach a goal.	Murphy [17]
...it should learn what it can to compensate for partial or incorrect prior knowledge	Russell and Norvig [18]
Autonomy refers to a robot's ability to accommodate variations in its environment. Different robots exhibit different degrees of autonomy; the degree of autonomy is often measured by relating the degree at which the environment can be varied to the mean time between failures and other factors indicative of the robots performance.	Thrun [19]
...agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal states.	Wooldridge and Jennings [20]

2.2.3 Trust Perspectives in Autonomous Operation

For the purposes of this work we define two perspectives on trust for autonomous systems: Design and Operational. These are summarised in Table 2.7. It is useful to further define and classify Operational Trust into two distinct but related sections defined in Table 2.8.

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.

We have already defined two trust perspectives when considering the design and operation of autonomous systems (Table 2.7). Examples of roles that interact with a

Work out how to reference across chapters in a multi doc

Table 2.5: Levels of Decision Making Automation (Sheridan and Verplank [21])

LOA	Description
1	The computer offers no assistance; the human must make all decisions and actions
2	The computer offers no assistance; the human must make all decisions and actions
3	The computer offers a complete set of decision/action alternatives, or
4	Narrows the selection down to a few, or
5	Suggests one alternative
6	Executes that suggestion if the human operator approves, or
7	Allows the human a restricted time to veto before automatic execution, or
8	Executes automatically, then necessarily informs the human, and
9	Informs the human only if asked, or
10	Informs the human only if it, the computer, decides to

system from both of these trust perspectives are provided in Tables 2.9 and 2.10.

Table 2.6: Levels of Automation (Endsley and Kaber [22])

LOA	Description
Manual Control	The human monitors, generates options, selects options (makes decisions), and physically carries out options.
Action Support	The automation assists the human with execution of selected action. The human does perform some control actions.
Batch Processing	The human generates and selects options; then they are turned over to automation to be carried out (e.g., cruise control in automobiles)
Shared Control	Both the human and the automation generate possible decision options. The human has control of selecting which options to implement; however, carrying out the options is a shared task.
Decision Support	The automation generates decision options that the human can select. Once an option is selected, the automation implements it.
Blended Decision Making	The automation generates an option, selects it, and executes it if the human consents. The human may approve of the option selected by the automation, select another, or generate another option.
Rigid System	The automation provides a set of options and the human has to select one of them. Once selected, the automation carries out the function.
Supervisory Control	The automation selects and carries out an option. The human can have input in the alternatives generated by the automation.
Automated Decision Making	The automation generates options, selects, and carries out a desired option. The human monitors the system and intervenes if needed (in which case the level of automation becomes Decision Support).
Full Automation	The system carries out all actions.

	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	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	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.

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. Given that systems that have high-levels of autonomy are designed to behave adaptively to dynamic environments, it is challenging to fully predict such non-deterministic behaviours prior to operational deployment. For example, in a navigation system it is difficult to predict the dynamic environment it will need to adapt to. Trust needs to be developed that the 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 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. Operational Trust is functionally derived from, but distinct from Design Trust.

Table 2.7: Trust Perspectives with respect to autonomous systems

Hard Trust or technical trust	The quantitative 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 and monitor the ability of a system, or another actor within a system, to perform a task unsupervised.
Soft Trust or common trust	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 human form of trust and is the main motivational driver for the human-factors trust discussion in *OTHER CHAPTER*. Can be rephrased as the level of confidence an operator has in an actor to perform a task unsupervised.

Table 2.8: Trust Perspectives within Operational Trust

	Role		
	Commander	**Operator**	**User**
Definition	Responsible for the system tactical activity (e.g. mission / activity setting)	Responsible for the ongoing 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**

TABLE 2.10: Examples of Roles that require a Operational Perspective of Trust in Autonomous Systems.

2.2.4 Design Trust

Five aspects of Design Trust have been identified: by whom

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, compliance with existing military and commercial standards can play a significant role in demonstrating the trustworthiness of any systems design. 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 context.

There are three main organisations that are developing or have developed assurance standards for Unmanned Systems in commercial, civil and military applications:

LOI	
1	Indirect receipt/transmission of UAV related payload data
2	Direct receipt of 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

TABLE 2.11: Levels of Interoperability for STANAG 4586 Compliant UCS

- 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[23] was promulgated in 2005 and defined a logistic and interoperability framework to provide commonality in the command and comtrol 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 Level of Interoperabilty (LOI) for compliant UAV systems, shown in Table 2.11. This structure has been criticised as being short sighted and at odds with the reality of modern and proposed autonomous vehicle operations [24], specifically that in modern autonomous systems, there is no such thing as direct control or Operator-in-the-loop, especially in the case of Beyond Line of Sight (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.

Further, 4586 predominantly deals with a one-to-one mapping between operators and nodes, 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 operating agent managing a group of autonomous nodes in-field, and handing off vehicle management responsibilities to the individual nodes.

SAE Levels of Autonomy possibly from [25]

Society of Automotive Engineers (SAE)

The AS-4 steering group is responsible for the development and maintenance of the Joint Architecture for Unmanned Systems (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[26].

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 Explosive Ordnance Disposal (EOD) family of robotic platforms.

American Society of Testing and Materials (ASTM)

The ASTM F38 committee has developed a Line of Sight (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[27]. However, the F41 Committee, focused on Unmanned Maritime Vehicle Systems (UMVSs) 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 [28], 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.

2.2.5 Human Factors related to Operational Trust

This work is considering autonomous systems as entities of wider systems, we refer to these here as Autonomous Collaborative Systems (ACSs). 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

[1] JAUS Service Interface Definition Language
[2] SAE AS6009, AS 6062, AS 6060, and AS 6057 respectively

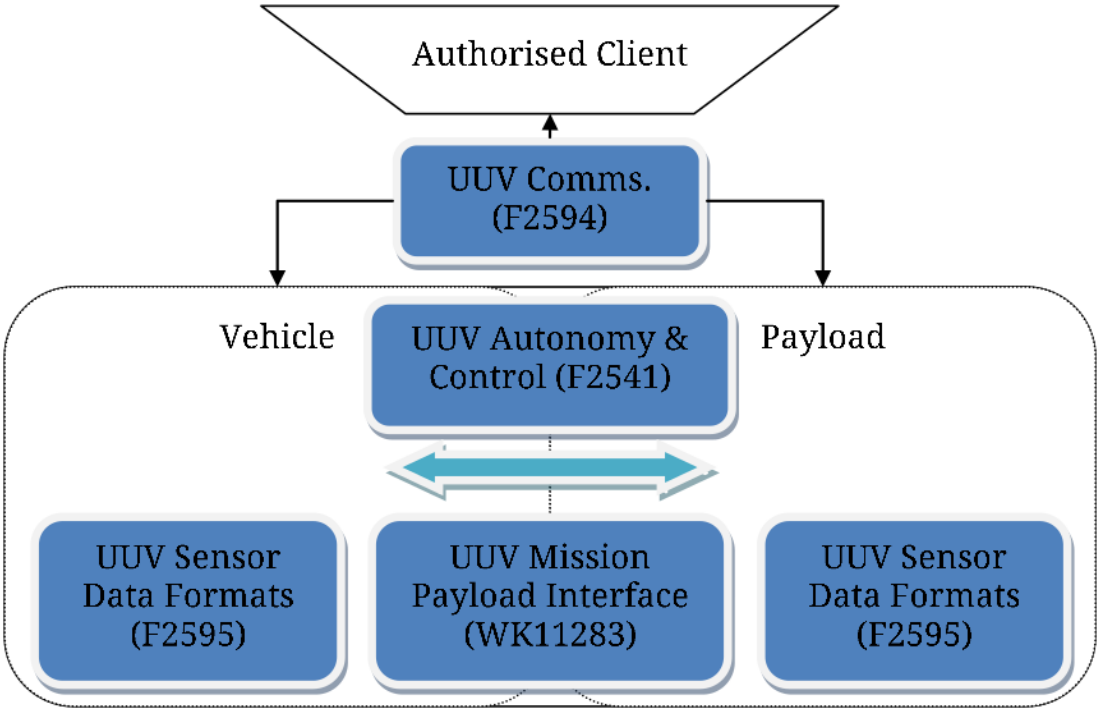


FIGURE 2.3: ASTM F41 UMVS Architecture (with relevant substandards in parenthesis)

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 HSC, summarised below:[24]

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 many domains. Particularly in defence and military applications, activity is characterised by long periods of "routine" punctuated by high intensity, usually unpredictable, activity. At those interfaces between "calm" and "storm", where real time situational awareness is imperative, temporary Information Overload is highly probable. Adaptive Automation enables autonomous systems to increase their 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 command and control. A simple example of this is shared watch-logs in Naval operations, 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 situational awareness 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.*

Check Security

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

In many areas, operators and commanders are required to make rapid decisions with imperfect information, driven by massively increased information availability and rates of change in areas such as battlefield tactics and global finance markets. However, Human decision making isnt always rational (especially under pressure), and operators use personally derived heuristics to make "rational shortcuts". 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 perspective) for any contradictory information, and will unquestionably follow, increasing error rates significantly.

This behaviour isn't 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

[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.

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 / 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.

Check Security

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 [29], 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 [30]. 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" perspective, 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.

2.2.6 Conclusions

ReDo this later

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[24].

There are opportunities 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[31],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[32] 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.

Need to check security status of this source

Need to check security status of this source

Need to check security status of this source

2.3 Trust in MANETs

2.3.1 Trust Model Design Considerations

From the previous sections, we define Trust as "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. There are five topics that are important to address in any MANETs trust model [33]:

- 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 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.

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

- 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 may attempt to subvert the model in order to get the unfairly good trust value.

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

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

2.3.2 Attacks on MANETs

2.3.3 Trust Management Frameworks

Distributed trust management frameworks for MANETs aim to detect, identify, and mitigate the impacts of malicious actors by distributing per-node assessments and opinions to collectively self-police behaviour.

This predictive behaviour attempts to solve four important problems (paraphrased from [12]):

- *Decision support* - For example; making informed routing table decisions based on past successes/failures.
- *Adaptability* - Ongoing prediciction of the networks future trust states directly detemines the risk faced by the network. Internalised knowledge of the expected risk can aid in selecting appropriate measures/ countermeasures such as automatically varying the level of authentication required for network activities.
- *Misbehaviour Detection* - Trust evaluation leads to a the natural policy that highly variable or low-trust nodes within a network should be subject to higher scrutiny; triggering this response indicates that a node is damaged or misbehaving.
- *Abstraction of Collective security characteristics* - Through per-node trust evaluation, the generalised trustworthiness of a set or subset of nodes can be derived to encapsulate the "health" of the network as a whole.

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;

- *Hermes Trust Establishment 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 [34].
- *Objective Trust Management Framework (OTMF)* takes a Bayesian approach and introduces the idea of applying a Beta function to changes in the per-link PLR over time, combining "Trust" and "Confidence of Assessment" into a single value [35]. Objective Trust Management Framework (OTMF) however does not appropriately combat multi-node-collusion in the network [36].
- *Trust-based Secure Routing [37]* 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)[36].

- *CONFIDANT*; [7] 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*; [38] 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.

- *Multi-parameter Trust Framework for MANETs (MTFM)* uses a number of communications metrics together for form a vector of trust, apply grey information theory to allow a system to detect and identify the tactics being used to undermine or subvert trust[39]

2.3.4 Single Metric Trust Frameworks

The Hermes trust establishment framework [34] uses Bayesian reasoning to generate a posterior distribution function of "belief", or trust, given a sequence of observations of that behaviour, $p(B|O)$(2.1).

Expand background detail on more frameworks

$$p(B|O) = \frac{p(O|B) \times p(B)}{\rho} \tag{2.1}$$

Where $p(B)$ is the prior probability density function for the expected normal behaviour, and ρ is a normalising factor.

Due to it's flexibility and simplicity, Hermes assumes that $p(B)$ is a Beta function, and therefore the evaluation of this trust assessment is based around the expectation value of the distribution (2.2) where α and β represent the number of successful and unsuccessful interactions respectively for a particular node i.

A secondary measurement of the confidence factor of the trust assessment t is generated as (2.3) and these measurements are combined to form a "trustworthiness" value T (2.4).

$$t_i \rightarrow E[\text{beta}(p|\alpha,\beta)] = \frac{\alpha_i}{\alpha_i + \beta_i} \tag{2.2}$$

$$c_i = 1 - \sqrt{\frac{12\alpha_i\beta_i}{(\alpha_i+\beta_i)^2(\alpha_i+\beta_i+1)}} \tag{2.3}$$

$$T_i = 1 - \frac{\sqrt{\frac{(t_i-1)^2}{x^2} + \frac{(c_i-1)^2}{y^2}}}{\sqrt{\frac{1}{x^2} + \frac{1}{y^2}}} \tag{2.4}$$

In (2.4), x and y are constants, used weight the two-dimensional polar mapping of trust and confidence assessments (t_i, c_i), and from [34], are taken as $x = \sqrt{2}, y = \sqrt{9}$.

Upon this per-node assessment methodology, OTMF overlays an observation distribution protocol so as to make the measurements α_i and β_i representative of the direct and 1-hop networks observations of the target node i, as well as expiring old observations from assessment and eliminating observations from "untrustworthy" nodes.

To date this work has been mostly limited to terrestrial, RF based networks. 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.

Want at least CONFIDANT and Fuzzy in here for contrast

Hermes, OTMF, CONFIDANT, and Fuzzy Trust-Based Filtering can be generalised as single-value probabilistic estimation, based on a Bayesian idea of taking a binary input state and generating an idealised Beta Distribution (2.5) of the future states of that input generated through an expectation value based on interactions (2.6).

$$\text{beta}(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}, \text{where } 0 \leq p \leq 1; \alpha,\beta > 0 \tag{2.5}$$

$$E(p) = \frac{\alpha}{\alpha+\beta} \tag{2.6}$$

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 is aware of the observed 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.

2.3.5 Multi-Metric Trust Frameworks

Given the potential incentives to a selfish attacker and potential threats to trust and fairness in sparse, noisy, and constrained environments, single metric trusts discussed above do not suitably cover the exposed threat surface.A multi-metric approach may be more

Probably best to just send a reference forward

appropriate to capture and monitor the realities of harsh and sparse communications environments.

MTFM[39] uses Grey Theory[40] to perform cohort based normalization of metrics at runtime, providing a "grey relational grade" of trust compared to other observed nodes in that interval for individual metrics, 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. This presents a stark difference between the Grey and Probabilistic approaches. Grey assessments are relative in both fairly and unfairly operating networks. All nodes will receive mid-range trust assessments if there are no malicious actors as there is nothing "bad" to compare against, and variations in assessment will be primarily driven by topological and environmental factors. Guo et al.[39] demonstrated the ability of grey relational analysis (GRA) to normalise and combine disparate traits of a communications link such as instantaneous throughput, received signal strength, etc. into a grey relational coefficient (GRC), or a "trust vector" in this instance.

The grey relational vector is given as

$$\theta^t_{k,j} = \frac{\min_k |a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|}{|a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|} \qquad (2.7)$$
$$\phi^t_{k,j} = \frac{\min_k |a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|}{|a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|}$$

where $a^t_{k,j}$ is the value of an 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 $\{a^t_{k,j} k = 1, 2 \dots K\}$, i.e. $g_j = \max_k(a^t_{k,j})$, $b_j = \min_k(a^t_{k,j})$ (where each metric is selected to be monotonically positive for trust assessment, e.g. higher throughput is presumed to be always better).

Weighting can be applied before generating a scalar value (3.16) allowing the detection and classification of misbehaviours.

$$[\theta^t_k, \phi^t_k] = \left[\sum_{j=0}^{M} h_j \theta^t_{k,j}, \sum_{j=0}^{M} h_j \phi^t_{k,j} \right] \qquad (2.8)$$

Where $H = [h_0 \dots h_M]$ is a metric weighting vector such that $\sum h_j = 1$, and in unweighted case, $H = [\frac{1}{M}, \frac{1}{M} \dots \frac{1}{M}]$. θ and ϕ are then scaled to $[0, 1]$ using the mapping $y = 1.5x - 0.5$. To minimise the uncertainties of belonging to either best (g) or worst (b) sequences in (3.15) the $[\theta, \phi]$ values are reduced into a scalar trust value by $T^t_k = (1 + (\phi^t_k)^2/(\theta^t_k)^2)^{-1}$ [41]. MTFM combines this GRA with a topology-aware weighting scheme (3.17) and a fuzzy whitenization model (3.18).

There are three classes of topological trust relationship used; Direct, Recommendation, and Indirect, repeating those discussed in section 2.1.3.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

This is currently half empty

relationships with both n_i and n_j, gives its assessment of n_j to n_i. In 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 node sets, N_R and N_I containing the nodes that have recommendation or indirect, relationships to the observing node respectively.

$$\begin{aligned} T_{i,j}^{MTFM} =& \frac{1}{2} \cdot \max_s\{f_s(T_{i,j})\}T_{i,j} \\ &+ \frac{1}{2}\frac{2|N_R|}{2|N_R| + |N_I|} \sum_{n \in N_R} \max_s\{f_s(T_{i,n})\}T_{i,n} \\ &+ \frac{1}{2}\frac{|N_I|}{2|N_R| + |N_I|} \sum_{n \in N_I} \max_s\{f_s(T_{i,n})\}T_{i,n} \end{aligned} \tag{2.9}$$

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:

$$\begin{aligned} f_1(x) &= -x + 1 \\ f_2(x) &= \begin{cases} 2x & \text{if } x \leq 0.5 \\ -2x + 2 & \text{if } x > 0.5 \end{cases} \\ f_3(x) &= x \end{aligned} \tag{2.10}$$

In the case of the terrestrial communications network used in [39], 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].

Guo et al. demonstrated that when compared against OTMF and Hermes trust assessment, MTFM provided increased variation in trust assessment over time, providing more information about the nodes' behaviours than packet delivery probability alone can.

2.4 Conclusion

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 to ensure their continued security, reliability, and performance. With demand for smaller, more decentralised MANET systems in a range of domains and applications, as well as a drive towards lower per-unit cost in all areas, TMFs are going to be increasingly applied to resource constrained applications, as the benefits and efficiencies they present are significant. For the purposes of this work, we are concerned with the analytical establishment of hard trust within a topologically dynamic network of autonomous actors. Beyond the constraints of the communications environment, knock on pressures in battery capacity, on-board processing, and locomotion 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. These increasingly decentralised applications present unique threats against trust management [42].

One area of application is the underwater marine environment, where extreme challenges to communications present themselves (propagation delays, frequency dependent attenuation, fast and slow fading, refractive multipath distortion, etc.). In addition to the communications challenges, other considerations such as command and control isolation, as well as power and locomotive limitations, drive towards the use of teams of smaller and cheaper Autonomous Underwater Vehicle (AUV) platforms. In underwater environments, communications is both sparse and noisy. Therefore the observations about the communications processes that are used to generate the trust metrics, occur much less frequently, with much greater error (noise) and delay than is experienced in terrestrial RF MANETS.

As such, the use of trust methods developed in the terrestrial MANET space should be reappraised for application within the underwater context [43].

In the next chapter, the marine communications environment will be studied, as will the current state of the art in the use of autonomy in specifically defence related maritime applications.

Chapter 3

Maritime Communications and Grey Theory

3.1 Maritime Communications Environment

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[44].

3.1.1 Mechanics of Acoustic Transmission

Unlike in RF energy energy transfer (where photons move through space to transmit energy from one place to another), acoustic wages are the result of mechanical pertubation of a medium where localised compressions and extensions pass energy across a medium through that mediums elastic properties. These "compression waves" propogate away from its source, and the rate of this propogation is the sound speed, velocity or c, measured in ms^{-1}. Acoustic pressure is usually measured in *Pascals* ($Pa/\mu Pa$). This is not to be confused with the fluid velocity corresponding to the instantenous motion of particles in the medium.

Hydrophones, like their more common microphone equivalent in air, are fundamentally pressure sensors. In the underwater environment, the dynamic range (difference between instantaneous high and low pressure values) may be extremely high, often more than 10 orders of magnitude higher. As such, logarithmic notation is justified.

Best to discuss notation here

Useful acoustic signals are generally maintained vibrations rather than instantenous pulses. They are characterised by their frequency f expressed in Hertz (Hz) or by their

Period (T, related to frequency by $T = 1/f$) In commonly used underwater acostics, used frequencies range from $\approx 10Hz - 100kHz$ depending on application.[45].

As with all waves, the relationship between frequency, period and the wavelength is given as in (3.1). As such the generally used upper and lower bounds of wavelength in most applications is from $1.5m@10Hz$ to $0.015m@100kHz$.

This wide range of frequencies and wavelengths allow for a diverse set of constraining factors; (Paraphrased from [46]).

- *Attenuation* in water; limiting the maxiumu usable range, which increases very rapibly with freqency
- *Dimensions* of soud source; which increase at lower f for a given transmission power
- *Spatial Selectivity* of sources and receivers as f increases, due to similarly increasing directivity of energy propogation.
- *Acoustic Reponse* of target surfaces (analogous to receiver gain in RF networks.

$$\lambda = cT = \frac{c}{f} \tag{3.1}$$

3.1.2 Velocity and density

Air has a baseline density of approximatly $1.3kgm^{-3}$, and the speed of sound is typically static around $340ms^{-1}$ In sea water, acostic wave velocity is close to $c = 1500ms^{-1}$ (generally between $1450ms^{-1}$ and $1550ms^{-1}$ depending on temperature, pressure, salinity etc.) Similarly variable is sea water denisty, which is nominally around $\rho = 1030kgm^{-3}$

While the sea/air surface is (ideally) a simple refractive interface, the interface between open seawater and marine sediment is gradiated, with density ranges between $1200 - 2000kgm_3$. This results in refractive and reflective velocities in the sediment interface ranging from $1500 - 2000ms^{-1}$.[46]

For comparison, the speed of light in air/water is $2.99\times10^8ms^{-1}$ and $2.249\times10^8ms^{-1}$.

this might be better as a table

Mackenzie proposed a more accurate model of acoustic velocity incorporating archival data from 15 worldwide sites that takes Temperature, Salinity and Depth into consideration [47]

$$\begin{aligned} c =& 1448.96 + 4.591T - 5.304 \times 10^{-2}T^2 + 2.374 \times 10^{-4}T^3 && (3.2)\\ &+ 1.340(S-35) + 1.630 \times 10^{-2}D + 1.675 \times 10^{-7}D^2 && (3.3)\\ &- 1.025 \times 10^{-2}T(S-25) - 7.139 \times 10^{-13}TD^3 && (3.4) \end{aligned}$$

Where T is the temperature in Celsius, S the salinity in parts per thousand, and D is the depth below the surface in meters.

$$10\log a(f) = 0.11 \cdot \frac{f^2}{1+f^2} + 44 \cdot \frac{f^2}{4100+f^2} + 2.75 \times 10^{-4} f^2 + 0.003 \tag{3.8}$$

FIGURE 3.1: Thorp's Absorption Model[45]

3.1.3 Intensity and Power

The energy of an acoustic wave is encapsulated into its kinetic and potential parts; where its kinetic energy corrsponds to the active motion energy of the particles in the medium, and the potential energy corresponding to the elastic potential of the medium in displacement/compression.

The acoustic intensity (I) is the energy flux mean value per unit of surface and time (3.5) in Watts/m^2 where p_0 is the plane wave amplitude (pressure) and $P_{rms} = p_0/\sqrt{2}$

$$I = \frac{p_0^2}{2\rho c} = \frac{p_{rms}^2}{\rho c} \tag{3.5}$$

3.1.4 Attenuation

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

$$A_{\text{aco}}(d, f) = A_0 d^k a(f)^d \tag{3.6}$$

$$10\log A_{\text{aco}}(d, f)/A_0 = k \cdot 10\log d + d \cdot 10\log a(f) \tag{3.7}$$

where A_0 is a unit-normalising constant, k is a geometric spreading factor (commonly taken as 1.5 for practical use, by may be 2 for true spherical propagation or 1 for plane-wave propagation)), and $a(f)$ is the absorption coefficient, that may be modelled in a variety of ways.

Thorp's formula (Equation 3.8) is very simple, only depending on f, and is designed to be most accurate about a temperature of 4°C at a depth of $\approx 1Km$. The Ainslie & McColm model is more complex, and incorporates the acidity of the water (H^+) as well as temperature (T), salinity (S in parts per trillion) but not depth Equation 3.10. The Fisher-Simmons model (**??**)is significantly more complex, taking into account the effects of boric acid concentrations and dissolved magnesium sulphate. While there are several limitations on this model in terms of its being fixed at a salinity of 35 ppt and a pH of 8, as this model incorporates depth, temperature, distance and frequency, it is very attractive for research directed at high variability environments.

Possibly need to switch this with the Francois Garrison model which, depend-ing on your

3.1.5 Ambient Noise Model

Ambient ocean noise can be assumed to be Gaussian with a continuous power spectral density in dB re μPa per Hz, driven by four major factors, shown in Table 3.1 [48].

$$
\begin{aligned}
10\log a(f) =& 0.106\frac{t_1 f^2}{t_1^2+f^2}e^{\frac{H^+-8}{0.56}} \\
&+ 0.52\left(1+\frac{T}{43}\right)\left(\frac{S}{35}\right)\frac{t_2 f^2}{t_2^2+f^2}e^{\frac{-D}{6}} \\
&+ 4.9\times 10^{-4} f^2 e^{-\left(\frac{T}{27}+\frac{D}{17}\right)}
\end{aligned} \tag{3.9}
$$

Where

$$
\begin{aligned}
t_1 =& 0.78\sqrt{\frac{S}{35}e^{\frac{T}{26}}} \\
t_2 =& 42e^{\frac{T}{17}}
\end{aligned}
$$

FIGURE 3.2: Ainslie & McColm Absorption Model

$$
10\log a(f) = A_1P_1\frac{t_1 f^2}{t_1^2+f^2} + A_2P_2\frac{t_2 f^2}{t_2^2+f^2} + A_3P_3f^2 \tag{3.10}
$$

Where

$$
\begin{aligned}
A_1 =& 1.03\times 10^{-8} + 2.36\times 10^{-10}\cdot T - 5.22\times 10^{-12}\cdot T^2 \\
A_2 =& 5.62\times 10^{-8} + 7.52\times 10^{-10}\cdot T \\
A_3 =& \left(55.9 - 2.39\cdot T + 4.77\times 10^{-2}\cdot T^2 - 3.48\times 10^{-4}\cdot T^3\right)\times 10^{-15} \\
t_1 =& 1.32\times 10^3\,(T+273.1)\,e^{\frac{-1700}{T+273.1}} \\
t_2 =& 1.55\times 10^7\,(T+273.1)\,e^{\frac{-3052}{T+273.1}} \\
P_1 =& 1 \\
P_2 =& 10.3\times 10^{-4}\cdot P + 3.7\times 10^{-7}\cdot P^2 \\
P_3 =& 3.84\times 10^{-4}\cdot P + 7.57\times 10^{-8}\cdot P^2
\end{aligned} \tag{3.11}
$$

FIGURE 3.3: Fisher-Simmons Absorption Model

TABLE 3.1: Contributing factors to Ocean Ambient Acoustic Noise

Source	Approximation
Turbulence	$10\log N_t(f) = 17 - 30\log f$
Shipping	$10\log N_s(f) = 40 + 20(s-0.5) + 26\log f - 60\log(f+0.03)$
Wind Driven Waves	$10\log N_w(f) = 50 + 7.5w^{\frac{1}{2}} + 20\log f - 40\log(f+0.4)$
Thermal Noise	$10\log N_{th}(f) = 15 + 20logf$

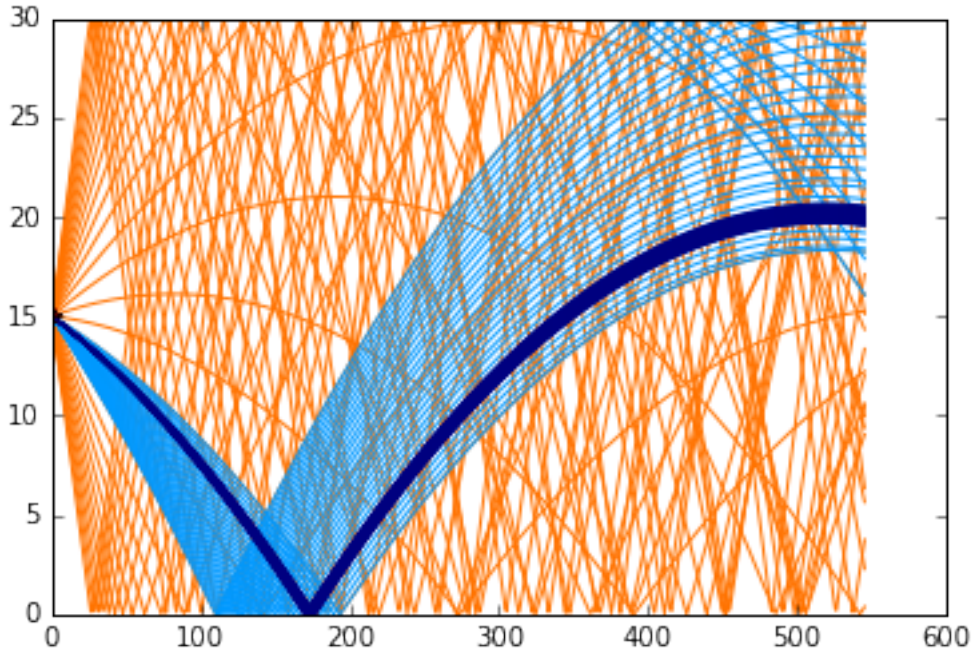


FIGURE 3.4: Non-Linear Marine Propagation in an isothermal profile

Vectorise and Label

3.1.6 Multipath effects

Refractive lensing and the multi-path nature of the medium result in line of sight propagation being extremely unreliable for estimating distances to targets. The first arriving acoustic signal has as the very least curved 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.

$$A_{\mathrm{RF}}(d,f) \approx \left(\frac{4\pi df}{c}\right)^2 \text{ where } c \approx 3\times 10^8 ms^{-1} \tag{3.12}$$

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

$$H(d,f) = \sum_{p=0}^{P-1} h(p) = \sum_{p=0}^{P-1} \Gamma_p / \sqrt{A(d_p,f)} e^{-j2\pi f \tau_p} \tag{3.13}$$
$$\text{where } \tau_p = d_p/c, c \approx 1500 ms^{-1}$$

where $d = d_0$ is the minimal path length between the transmitter and receiver, $d_p, p = \{1, \ldots P-1\}$ 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 terrestrial RF ($A_{\mathrm{aco}} \propto f^{2d}$ vs $A_{\mathrm{RF}} \propto (df)^2$). While both frequency dependant factors are quadratic, approximating the factors in (3.8), $f \propto A_{\mathrm{aco}}$ is at least 4 orders of magnitude higher than $f \propto A_{\mathrm{RF}}$

3.1.7 Modelling and Simulation of the Acoustic Medium / Channel

Several toolkits exist in a variety of states that perform communications agent simulation, most notably the NS-2 / 3 family of frameworks and their addons. Some of these frameworks, such as SUNSET [49] and AquaTools [**?**]

3.1.8 Routing and Network Design for UANs

Forward Error Correction coding is used on such channels to minimise packet losses.

3.1.9 Need for Trust in Maritime Networks

As Autonomous Underwater Vehicle (AUV) platforms 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 [42]. 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.

Possibly worth having some discussion on mobility in here

3.2 Grey System Theory and Grey Trust Assessment

This section largely repeats from MTFM discussion but the maths needs explored somewhere

3.2.1 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.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;

$$\oplus_1 + \oplus_2 \in [a_1 + a_2, b_1 + b_2] \tag{3.14a}$$

$$-\oplus \in [-b, -a] \tag{3.14b}$$

$$\oplus_1 - \oplus_2 = \oplus_1 + (-\oplus) \tag{3.14c}$$

$$\oplus_1 \times \oplus_2 \in [\min(a_1a_2, a_1b_2, b_1a_2, b_2a_2), \max(a_1a_2, a_1b_2, b_1a_2, b_2a_2)] \tag{3.14d}$$

$$\oplus^{-1} \in [b^{-1}, a^{-1}] \tag{3.14e}$$

$$\oplus_1/\oplus_2 = \oplus_1 \times \oplus_2^{-1} \tag{3.14f}$$

$$\oplus \times k \in [ka, kb] \tag{3.14g}$$

$$\oplus^k \in [a^k, b^k] \tag{3.14h}$$

where k is a scalar quantity.

3.2.2 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. [50] 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.

3.2.3 Grey Sequence Buffers and Generators

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".

eqs of sequence buffers and partial derivs

3.2.4 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 entities in that interval, while maintaining the ability to reduce trust values 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 cohorts. Entities will receive mid-range trust assessments if there are no malicious actors as there is no-one else "bad" to compare against.

Guo[39] demonstrated the ability of Grey Relational Analysis (GRA)[40] 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 [39], 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

$$\begin{aligned} \theta^t_{k,j} &= \frac{\min_k |a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|}{|a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|} \\ \phi^t_{k,j} &= \frac{\min_k |a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|}{|a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|} \end{aligned} \tag{3.15}$$

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.

$$[\theta^t_k, \phi^t_k] = \left[\sum_{j=0}^{M} h_j \theta^t_{k,j}, \sum_{j=0}^{M} h_j \phi^t_{k,j} \right] \tag{3.16}$$

Where $H = [h_0 \ldots h_M]$ is a metric weighting vector such that $\sum h_j = 1$, and in the basic case, $H = [\frac{1}{M}, \frac{1}{M} \ldots \frac{1}{M}]$ 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 (3.15).

MTFM combines this GRA with a topology-aware weighting scheme(3.17) and a fuzzy whitenization model(3.18). 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.

$$T_{i,j}^{MTFM} = \frac{1}{2} \cdot \max_s\{f_s(T_{i,j})\}T_{i,j} + \frac{1}{2}\frac{2|N_R|}{2|N_R| + |N_I|} \sum_{n \in N_R} \max_s\{f_s(T_{i,n})\}T_{i,n} + \frac{1}{2}\frac{|N_I|}{2|N_R| + |N_I|} \sum_{n \in N_I} \max_s\{f_s(T_{i,n})\}T_{i,n} \tag{3.17}$$

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:

$$\begin{aligned} f_1(x) &= -x + 1 \\ f_2(x) &= \begin{cases} 2x & \text{if } x \leq 0.5 \\ -2x + 2 & \text{if } x > 0.5 \end{cases} \\ f_3(x) &= x \end{aligned} \tag{3.18}$$

Grey System Theory, by it's own authors admission, hasn't taken root in it's originally intended area of system modelling [50]. 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.

Chapter 4

Assessment of TMF Performance in Marine Environments

4.1 Introduction

In this chapter, we demonstrate the need for multi-metric trust assessment in UAN.

In underwater environments, communications is both sparse and noisy. Therefore the observations about the communications processes that are used to generate the trust metrics, occur much less frequently, with much greater error (noise) and delay than is experienced in terrestrial RF MANETs. In addition to the communications challenges, other considerations such as command and control isolation, as well as power and locomotive limitations, drive towards the use of teams of smaller and cheaper AUVs. As such, the use of trust methods developed in the terrestrial MANET space must be reappraised for application within the underwater context [43]. Many UANs use MANET architectures, however the marine environment presents new challenges for trust management frameworks that have been developed for use in conventional (i.e. Terrestrial RF) MANETs. Previous research has established the advantages of implementing TMFs in 802.11 based MANETs, particularly in terms of preventing selfish operation in collaborative systems [6], and maintaining throughput in the presence of malicious actors [7]

this makes no sense

To date this work has been limited to terrestrial, RF based networks.

We investigate the operation of a selection of traditional MANET TMFs in this environment. We characterise these challenges and present results that demonstrate a multi-metric approach to Trust greatly enhances the effectiveness of TMFs in these environments.

In Section 4.2 we establish an experimental configuration for the marine space, and review the scenarios and results presented in [39]. In Section 6.2 we present our findings in trust establishment and malicious behaviour detection, comparing with other current TMFs (Hermes and OTMF) and analyse the use of this multi-parameter approach to detecting malicious and selfish behaviour in autonomous marine networks.

The contributions of this chapter are a study on the comparative operation and performance of TMFs in marine acoustic networks, and a review of metric suitability for TMFs in marine environments, informing future metric selection for experimenters and theorists. We also show that single metric trust systems are not directly suitable for the marine context in terms of the different threat and cost scenario in that environment. Finally, we demonstrate a methodology to assess the usefulness of metrics in discriminating against misbehaviours in such constrained, delay-tolerant networks.

Key parts of this chapter were presented at TrustCom-BigDataSE-ISPA 2015 as "Single and Multi-Metric Trust Management Frameworks for use in Underwater Autonomous Networks."[51]

These single metric TMFs provide malicious actors with a significant advantage if their activity does not impact 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. This causes a significant negative effect on 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.

For the purposes of this work, from those TMFs discussed in subsection 2.3.3, we select Hermes trust establishment, OTMF and MTFM as indicative single and multi metrics frameworks for comparison, as Hermes captures the core operation of a pure single metric assessment methodology and OTMF provides a comparison that combines assessments from across nodes to develop trust opinions.

From the discussion on the nature of the communications environment in section 3.1, it's clear that before assessing communications metrics a simulated underwater environment, appropriate scaling factors must be found that are realistic from an application perspective but are also comperable in some form to the MANET case.

4.2 System Model Characterization

4.2.1 Mobility, Topology, and Communications

We apply two mobility patterns for investigation; all nodes static and all nodes mobile. The reason for this is that in other mobility combinations, the node targeted for misbehaviour (n_1) will already be behaving differently compared to the rest of the network regardless of the misbehaviour.

The six nodes are initially arranged as per Fig. 4.1 with each node on average 100m from each other as per [39]. The use of six nodes and the particular layout enables the investigation of the three trust relationships based on minimum path topologies, such that the node generating the trust assessments, n_0 has Direct, Recommendation,

expand this section to include discussion and results of single mobility models

and Indirect trust assessments of n_1 available to it from itself, $[n_2, n_3]$, and $[n_4, n_5]$ respectively. (See Section 2.1.3)

Collaborations with NATO Centre for Maritime Research and Experimentation (CMRE) in La Spezia, and Defence Science and Technology Laboratorys (DSTLs) Naval Systems Group inform that this is a practical team-size for environmental and defence applications.

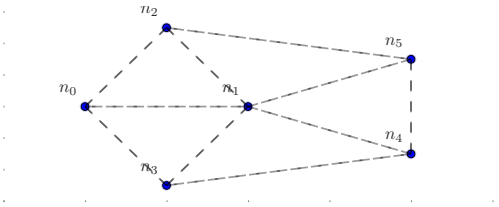


FIGURE 4.1: Initial layout with nodes spaced an average of 100m apart

4.2.2 Simulation Background

Simulations were conducted using a Python based simulation framework, SimPy [52], with a network stack built upon AUVNetSim [53], with transmission parameters (Table 4.1) taken from and validated against [45], [54] and [**?**]

Given the differences in delay and propagation between RF and marine networks, it would not be expected that the same application rates (e.g. packet emission rates or throughput) and node separations are equally stable in this environment. Therefore, we first characterise a zone of performance within which the network have stable operation.

it would be worth while going through this verification explicitly as an appendix

4.2.3 Scaling Considerations between Terrestrial and Underwater Environments

We establish an appropriate safe operating zone for marine communications by looking at the communications rate and physical distribution factors across the two selected mobility scenarios. From Table 4.1, the operating transmission range of this model of acoustic communications is ≈ 6 times further than that of 802.11, indicating that a suitable operating environment will have an area $\approx \sqrt{6}$ times the area of the 802.11 case. However, it was recognised in Section **??** that underwater, the relationship between attenuation and distance is exponential, so this would represent an upper bound of performance, where nodes are approximately $400m$ apart.

Exploratory simulations were run to further constrain this bound. As the separation is increased, the emission rate at which the network becomes saturated decreases, reducing overall throughput. This throughput degradation is tightly coupled with the mobility, as increasing mobility leads to increasing delays as routes are constantly broken, re-advertised and re-established. For instance, where all nodes are static, we do not see significant drops in saturation rates until node separation approaches 800m, nearly

TABLE 4.1: Comparison of system model constraints as applied between Terrestrial and Marine communications

Parameter	Unit	Terrestrial	Marine
Simulated Duration	s	300	18000
Trust Sampling Period	s	1	600
Simulated Area	km^2	0.7	0.7-4
Transmission Range	km	0.25	1.5
Physical Layer		RF(802.11)	Acoustic
Propagation Speed	m/s	3×10^8	1490
Center Frequency	Hz	2.6×10^9	2×10^4
Bandwidth	Hz	22×10^6	1×10^4
MAC Type		CSMA/DCF	CSMA/CA
Routing Protocol		DSDV	FBR
Max Speed	ms^{-1}	5	1.5
Max Data Rate	bps	5×10^6	≈ 240
Packet Size	bits	4096	9600
Single Transmission Duration	s	10	32
Single Transmission Size	bits	10^7	9600

double the initial estimate. When all nodes are randomly walking the saturation point collapses from 0.025pps at 300m to 0.015pps at 400m. Our results indicate that the best area to continue operating in for a range of node separations is at 0.015pps, and that a reasonable position scaling is from 100m to $300m$, beyond which communication becomes increasingly unstable, especially in terms of end-to-end delay. These results are similar to work performed in [53], and are expected in such a sparse, noisy, and contentious environment.

4.3 Establishing Scale Factors in Communications Rate

In this section we characterise the simulated communications environment, establishing an optimal packet emission rate for comparison against [39].

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. 4.2 and 4.3, 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. 4.3, 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 4.2: Varying packet emission rate demonstrates maximal throughput at 0.025 packets per second, equivalent to ≈240 bps

FIGURE 4.3: Varying packet emission rate demonstrates a saturation point at 0.025 packets per second

4.3.1 Establishing Scale Factors in Physical Distribution

In this section we characterise the effect of node-separation scaling on communications operation for comparison against [39]. 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 4.1, the operating transmission range of acoustic is ≈ 6 times further than 802.11, indicating that a suitable operating environment will have an area ≈ $\sqrt{6}$ 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. 4.4. This collision rate is well within the tolerances of the MAC layer, as shown in Fig. 4.5, where even with a rising collision rate, packets are being reliably received.

FIGURE 4.4: Comparison of Medium Acquisition Collisions, Throughput, and Enqueued packets against varying application packet emission rates.

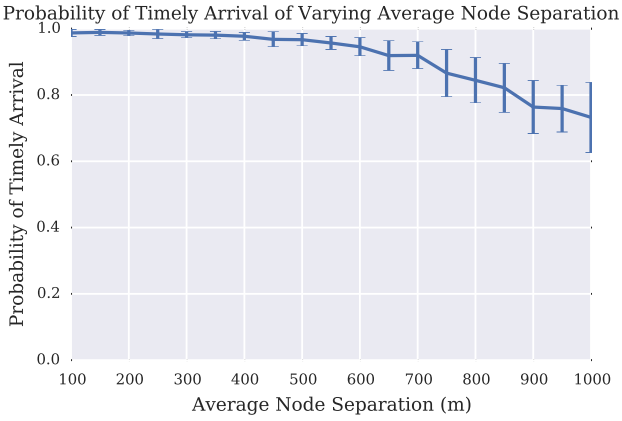


FIGURE 4.5: Probability of Timely Reception across a range of node scaling.

However, when end-to-end delay is investigated, it's clear from Fig. 4.6 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. 4.7.

According to Xu [55], 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. 4.7, 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.

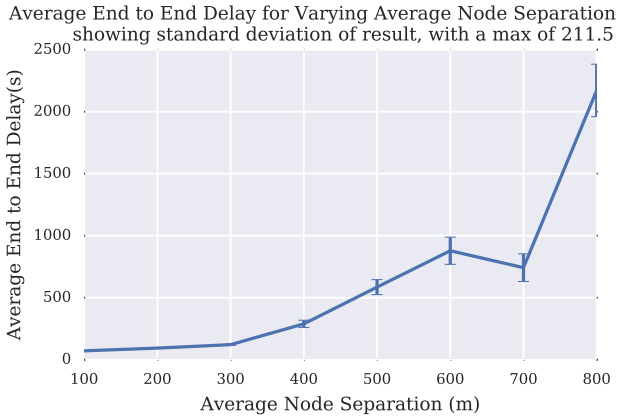


FIGURE 4.6: End to End Delay under varying node-separations

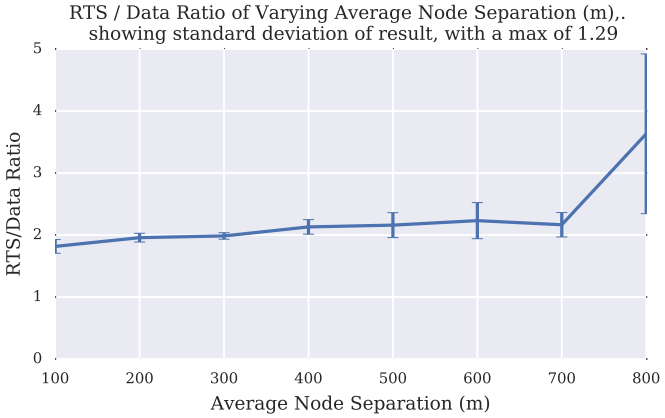


FIGURE 4.7: RTS/Data ratio for varying node-separations

TABLE 4.2: Tabular view of data from Figs 4.5, 4.6, and 4.7

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

Chapter 5

Strategies for Multi-Domain Trust Assessment

Chapter 6

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

6.1 Introduction

6.1.1 Selected Misbehaviours

We are primarily concerned with the direct trust relationship between n_0 and n_1, i.e. n_0's assessment of the trustworthiness of n_1, or $T_{1,0}$.

Guo et al. introduce a range of misbehaviours, including modification of the packet loss rate of routing nodes and limiting throughput on a per-link basis as well as a selection of combined misbehaviours. Given that the established links are already heavily constrained, such attacks would severely impact the general performance of the network beyond the scope of simple selfishness. These direct malicious behaviours effectively trigger saturation collapses in operating regions of the network that should be stable.

Therefore, we apply two more subtle misbehaviours to investigate;

1. Malicious Power Control (MPC), where n_1 increases its transmit and forwarding power by 20% for all nodes *except* communications from n_0 in order to make n_0 appear to be selfishly conserving energy to the rest of the team, while n_1 itself appears to be performing very well.

2. Selfish Target Selection (STS), where n_1 preferentially communicates, forwards and advertises to nodes that are physically close to it in effort to reduce its own power consumption.

6.2 Simulation Results and Discussion

Having established a safe operating range for comparison at 300m average separation and an emission rate of 0.015pps, we perform each of the three selected behaviours (Fair, MPC, STS) in both the static and mobile scenarios. We select a trust assessment period of 10 mins for a 5 hour mission to scale in comparison to relative bitrates experienced (1Mbps vs $\approx$ 15bps).

The six metrics used for grey assessment are; transmitted and received throughput and power, delay, and packet loss rate (PLR) as calculated by aborted and unacknowledged, transmissions. Compared to [39], this metric set lacks a data rate quantity as the network is not dynamically adjusting bandwidth. In context of GRC generation (3.15), the best sequence g was selected using the lowest PLR, delay, and powers, and the highest throughputs, and the worst sequence, b the inverse of these metrics, reflecting the observations made in Section 3.1.

The particular factors under discussion are the relative performance of MTFM against OTMF and Beta with respect to statistical stability across mobilities and in responsiveness to changing network behaviour. We establish a similar result set by initially tracking the resultant trust values established by MTFM in the pair of mobility scenarios, shown in Fig. A.2. We are also concerned with the opinions of n_1 provided to n_0 by other nodes, where $[T_{1,2}, T_{1,3}]$ and $[T_{1,4}, T_{1,5}]$ denote the sets of recommendation and indirect trust assessment respectively.

We also include aggregate assessments; $T_{1,\text{Avg}}$, the unweighted mean of direct trust assessments of n_1 from all nodes and $T_{1,\text{MTFM}}$, the final MTFM trust assessment value based on both network topology and whitenization from (3.18).

The variability in assessment is coupled to mobility; in the static case (Fig. 6.1a), we see that the nodes exhibit relatively consistent distributions. In the full mobility case, shown in Fig. 6.1b, this subjective variability is greatly increased. As the topology is highly dynamic, delays due to re-establishing routes can be very large, perturbing the trust value. The $T_{1,\text{MTFM}}$ displays a significantly reduced variation than those of the individual subjective observations in all cases, even when compared to the unweighted average, $T_{1,\text{Avg}}$. This demonstrates T_{MTFM}'s value as an aggregating trust assessment in such sparse and noisy environments. Further, in Fig. 6.1d we observe a much higher variability in assessment in T_0, correctly indicating that there is something wrong with the relationship between n_0 and n_1.

6.2.1 Comparison between MTFM, Hermes and OTFM

As per [39], "fair" scenarios were also performed with no malicious behaviour, applying OTMF and Hermes assessment as well as MTFM, providing like-for-like comparison of assessment. For simplicity of presentation, we only consider the fully-mobile scenario, as we are concerned with the establishment of trust in mobile networks

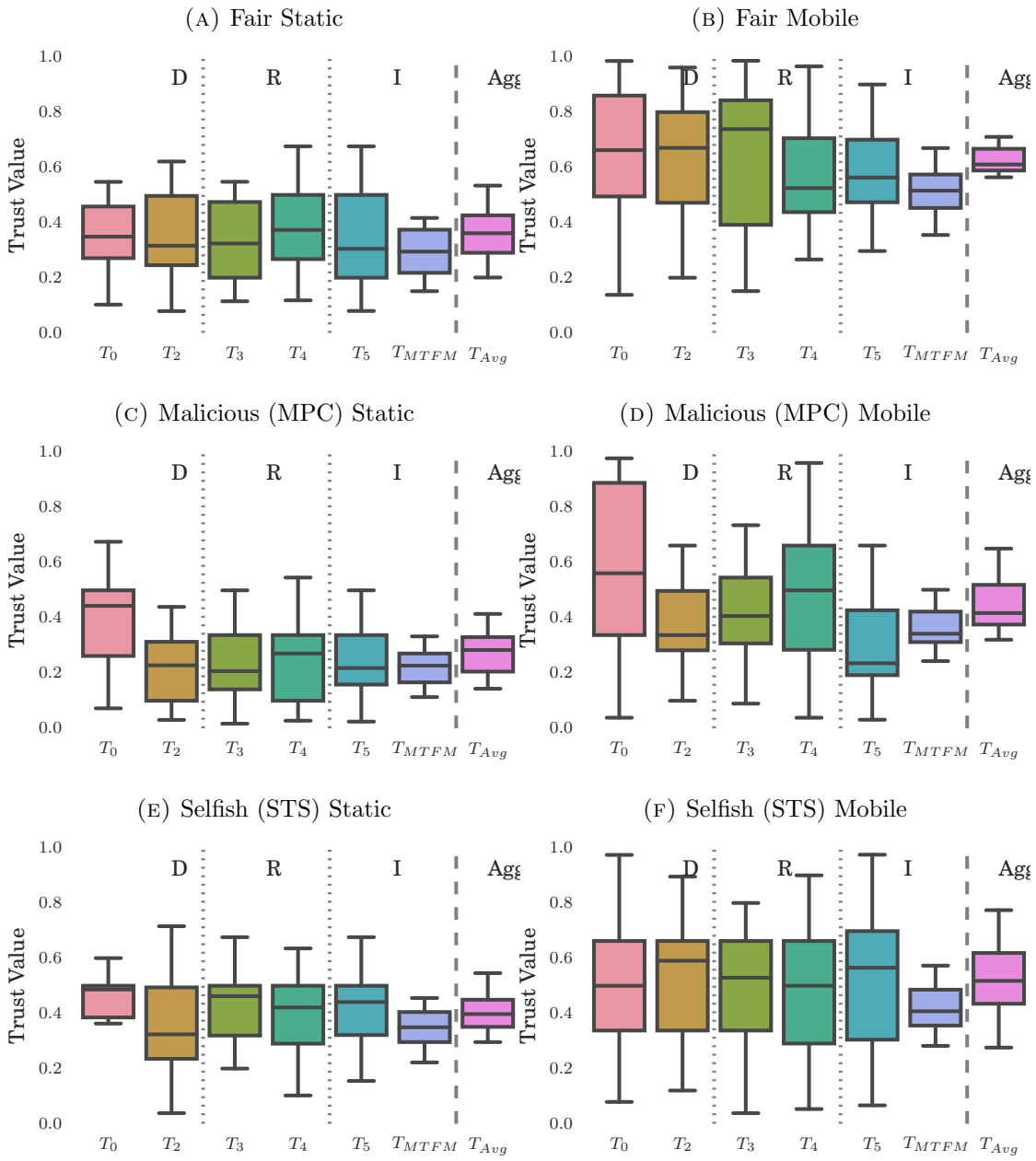
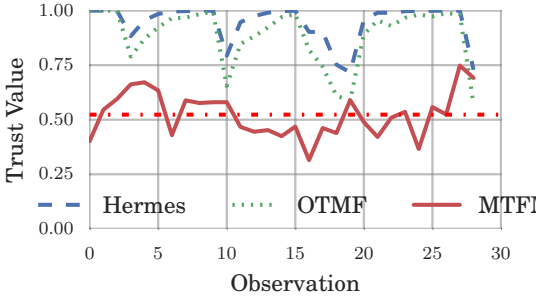
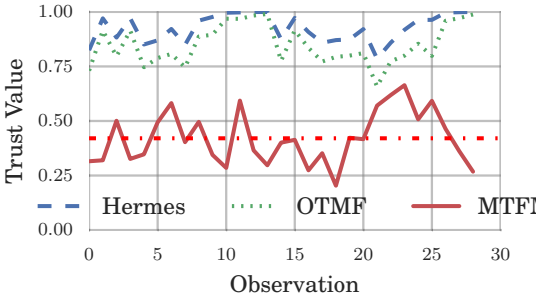


FIGURE 6.1: MTFM Trust assessments of n_1 ($T_{1,X}$), showing Direct, Recommender and Indirect relationships, as well as the Aggregate trust assessments from combining these

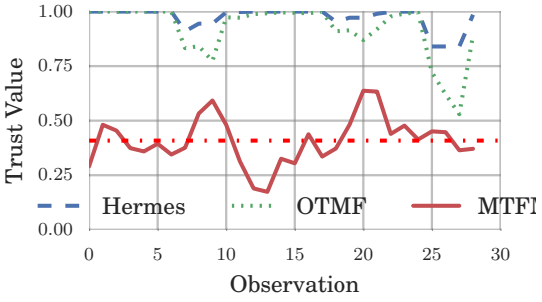
The use of Forward Beam Routing and a CSMA/CA MAC scheme from AUVNetSim [53] in our simulation mitigates a significant number of packet losses through collision avoidance and contention handling, leading to the situation that the only genuinely lost packets occur when a node moves completely out of range of any other node and time out occurs in route discovery rather than transmission. As such, confirmed packet losses are relatively rare and in a delaying network like this, it is difficult to set a differentiating time out between packets that are in the network but queued, and packets that are actually "lost".



(A) Fair Scenario



(B) Malicious Power Control (MPC) Scenario



(C) Selfish Target Selection (STS) Scenario

FIGURE 6.2: $T_{1,0}$ for Hermes, OTMF and MTFM assessment values for fair and malicious behaviours in the fully mobile scenario (mean of MTFM also shown)

The single metric TMFs used in conventional MANETs require regular and constant input to shape and adjust their evaluations, which for a network with significant and irregular delays such as this, is not practical. This renders OTMF and Hermes assessment at best uninformative and at worst misleading; consistently providing nodes a high trust assessment as they have very little information to extract trust from.

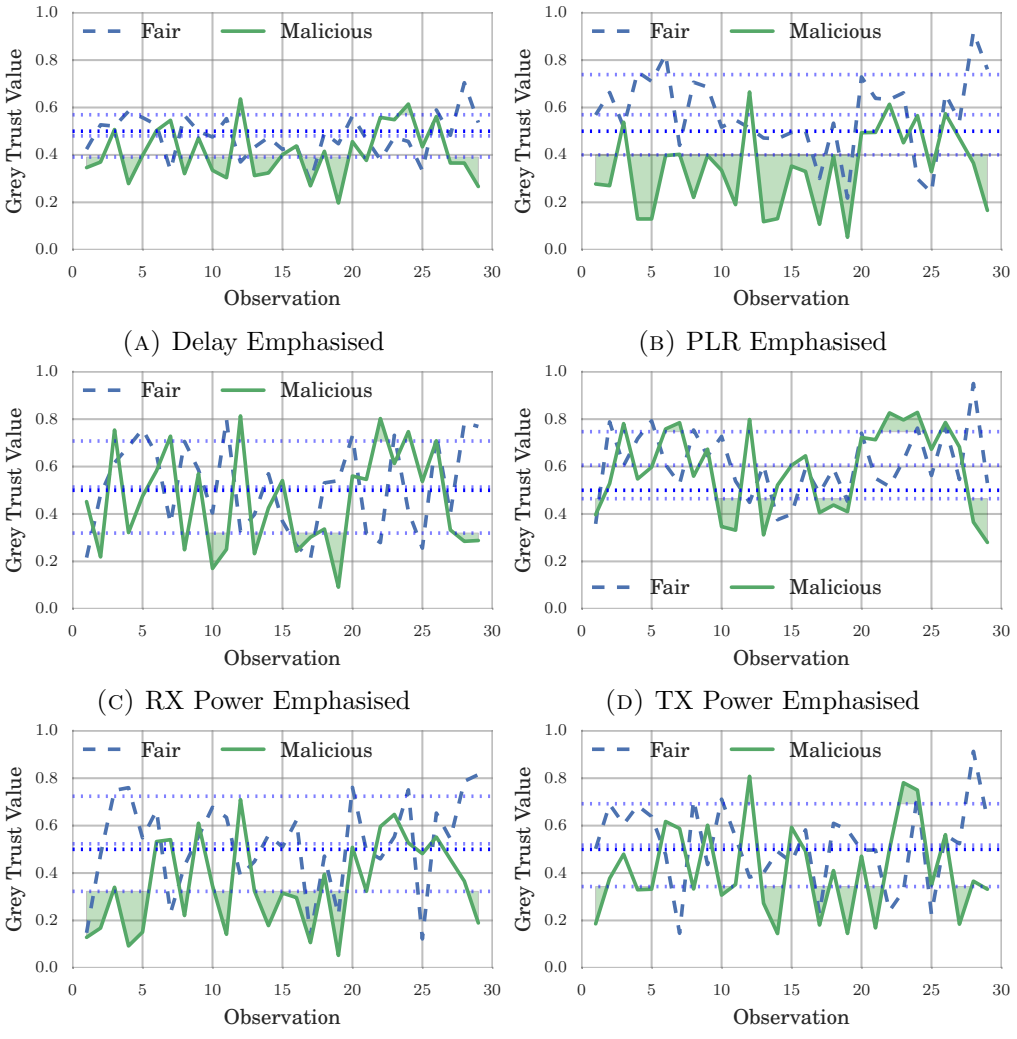
Fig. 6.2 shows a comparison between the unweighted response of MTFM compared to OTMF and Hermes assessment functions on the same data for the fair, malicious and selfish behaviours respectively. It is important to note a distinction between the expectations of MTFM compared to other TMFs; MTFM is primarily concerned with the identification of differences in the behaviours of nodes in a network, and is relative rather than absolute. That is to say that under MTFM, nodes are compared against the worst current performances across metrics of other observed nodes and graded against them, rather than the absolute (objective) approach taken by many TMFs. In these cases, particularly since the methods of attack were not directly related to PLR, OTMF and Hermes have not registered significant activity in either misbehaviour when compared to the fair scenario. The difference between the MTFM trust assessments under "fair" and "malicious" behaviour is lowered by $\approx 10\%$ in both cases, in terms of the mean values returned. At run time, similar results could be attained by an exponentially weighted moving average filter (EWMA).

On their own, neither OTMF, Hermes, or unbiased MTFM appear to be effective

in detecting or identifying malicious behaviour in this environment, in fact OTMF and Hermes don't appear to differentiate between fair and selfish scenarios at all.

6.2.2 Metric Weighting

We apply a sequence of vectors that preferentially weight each metric in Eq. (3.16) to each of the three simulation runs. For a metric weight vector H, where the metric m_j is emphasised as being twice as important as the other metrics, we form an initial weighting vector $H' = [h_i \dots h_M]$ such that $h_i = 1 \forall i \neq j; h_j = 2$. We then scale that vector H' such that $\sum H = 1$ by $H = \frac{H'}{\sum H'}$. Using this process we can extract and highlight the primary aspects of an attack by comparing against the deviation from the "fair" result set.



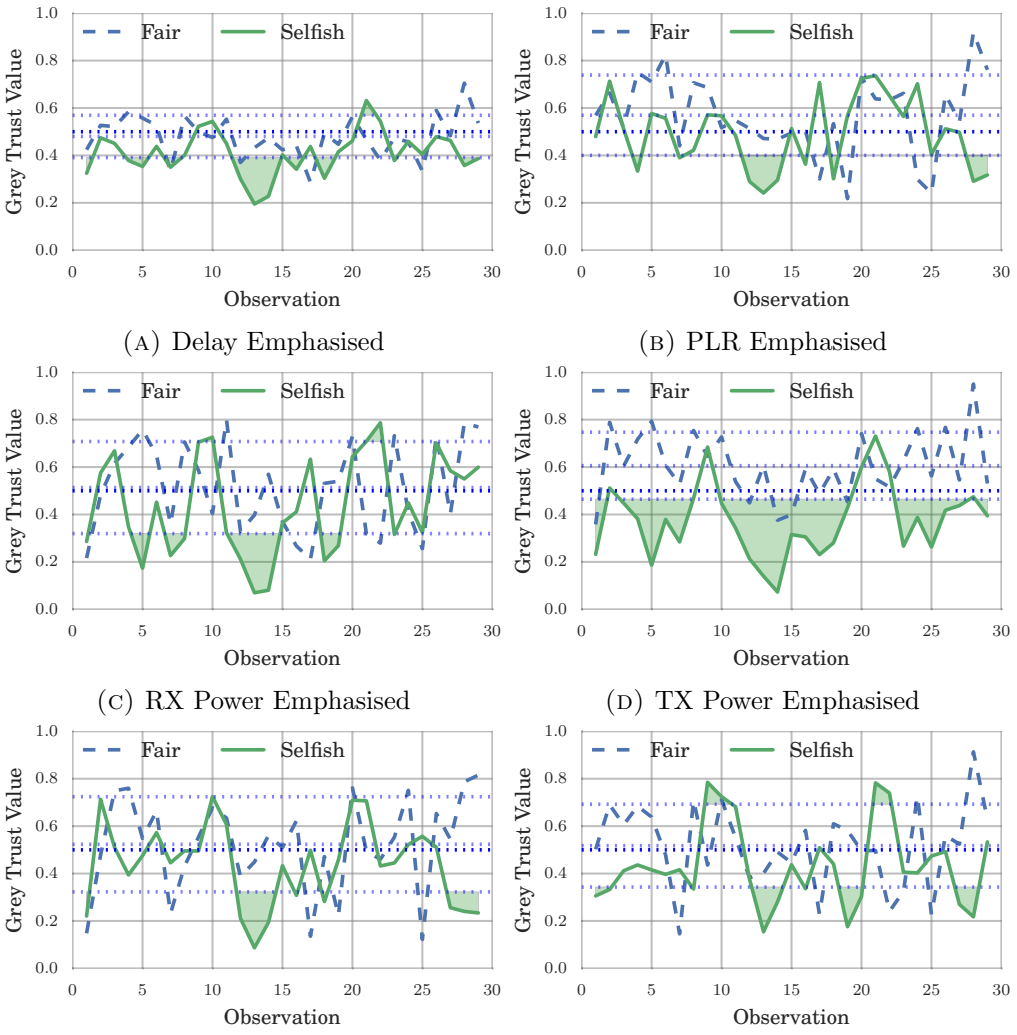
(A) Delay Emphasised (B) PLR Emphasised

(C) RX Power Emphasised (D) TX Power Emphasised

(E) RX Throughput Emphasised (F) TX Throughput Emphasised

FIGURE 6.3: $T_{1,MTFM}$ in the All Mobile case for the Malicious Power Control behaviour, including dashed $\pm\sigma$ envelope about the fair scenario

From Fig. 6.3 we can see that the malicious node is consistently outside the $\pm\sigma$ (one standard deviation above and below the mean) envelope of the fair scenario it's



(A) Delay Emphasised (B) PLR Emphasised

(C) RX Power Emphasised (D) TX Power Emphasised

(E) RX Throughput Emphasised (F) TX Throughput Emphasised

FIGURE 6.4: $T_{1,MTFM}$ in the All Mobile case for the Selfish Target Selection behaviour, including dashed $\pm\sigma$ envelope about the fair scenario

being compared to. This is particularly true for PLR, with smaller impacts on delay, received power and transmitted throughput. This weighted delta in received throughput is minimal to insignificant compared to the width of the detection envelope, occasionally breaching the envelope for a short period.

In the selfish case (Fig. 6.4) we observe much lower weighted delta in PLR and delay, with greatly increased impact on transmission power. In comparison to [39], these results are qualitatively similar, however here the differences between the fair case and the misbehaviours are less clear than in the comparable terrestrial space. Guo et al. show similar types of behaviour but report a weighted delta from ≈ 0.4 to ≈ 0.9 across the simulation period, compared to our maximum delta in P_{TX} in selfish behaviour (Fig. 6.4d) of ≈ 0.3 for an inconsistent interval.

6.2.3 Weight Significance Analysis for Behaviour Classification

For a more quantitative assessment of the viability of multi-metric trust assessment methods, we take the qualitative analysis above and apply a Random Forest regression [56] to assess the relative importance of the selected metrics on relative detectability of malicious behaviour. Random Forest accomplishes this by generating a large number of random regression trees and prune these trees to fit incoming data. The target function for this regression was the area between the target behaviours weighted T_{MTFM} curve and the $\pm\sigma$ envelope of the base behaviour as shaded in Figs. 6.3 and 6.4. From this training process we can extract the relative importance of each input feature (metric) in terms of how good it is to differentiate between the fair case and a given misbehaviour. Additionally we perform a cross correlation analysis to establish the correlations between given metric weighting emphasis and the output of the target function. Our intention is to establish the metrics that not only differentiate both misbehaviours from the fair case, but also what metrics differentiate the two misbehaviours from each other.

Applying this target regression to 729 different metric weight vector emphasis combinations reveals that each of the three combinations (i.e. comparing fair to misbehaviours, and comparing the misbehaviours) present distinct patterns of significance in three primary metrics; received throughput, transmitted power, and PLR, with delay, received power and transmitted throughput playing a lesser role. Practically this means that in order to accurately distinguish between these scenarios, these primary metrics should be higher-weighted in the generation of $T_{1,MTFM}$ in (3.17).

It may initially appear odd that the relative significance of the received throughput is similar between all three scenario combinations, however a correlation analysis shows that in the MPC attack; the received throughput is positively correlated with successful classification against the fair case ($R = +0.71, p \approx 10^{-100}$), while the inverse is the case for the STS attack ($R = -0.70, p \approx 10^{-100}$). It is expected that Transmitted power should be the defining characteristic of STS ($R = +0.72, p < 10^{-100}$) as the node is acting fairly from a protocol perspective but is acting unfairly at a higher (incentive) level; it is performing fairly in terms of it's communications with other nodes, however it is preferring to communicate with nodes that it can expend less energy communicating with. A summary of these correlations is shown in Table. 6.1.

Comparing Figs. 6.2, 6.3b, and 6.4b, while it is possible that in a cleaner, less sparse, and less noisy environment, OTMF would be able to detect the MPC behaviour, from Fig. 6.5 we see that PLR plays almost no part at all in detecting the STS behaviour, and so OTMF would not detect the attack.

As such this presents the open opportunity to develop a heuristic weight search scheme to detect malicious behaviour without the comparison to the fair scenario. This would be accomplished by assessing the impact of differential metric weighting on the mean trust assessment rather than comparing co-weighted valuations across scenarios.

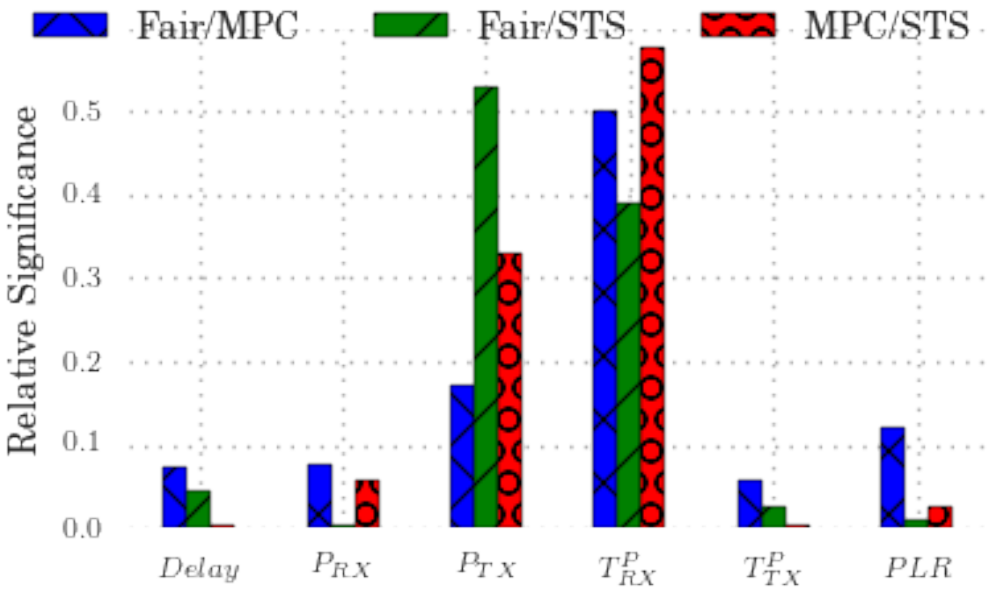


FIGURE 6.5: Random Forest Factor Analysis of Malicious (MPC), Selfish (STS) and Fair behaviours compared against eachother

TABLE 6.1: Correlation Coefficients between metric weights and behaviour detection targets

Correlation	Delay	P_{RX}	P_{TX}	T^P_{RX}	T^P_{TX}	PLR
Fair / MPC	0.199	0.159	-0.416	0.708	-0.238	-0.401
Fair / STS	0.179	-0.009	0.724	-0.697	-0.145	-0.052
MPC / STS	0.058	-0.134	0.146	-0.768	0.052	0.146

6.3 Conclusions and Future Work

We have demonstrated that existing MANET Trust Management Frameworks are not directly suitable to the sparse, noisy, and dynamic underwater medium. We presented a comparison between trust establishment in MANETs in a simulated underwater environment, demonstrating that in order to have any reasonable expectation of performance, throughput and delay responses must be characterised before implementing trust in such environments. While the MTFM value does not display any immediate difference between the two behaviours, we have shown that by exploring the metric space by weight variation, the existence and nature of the malicious behaviour can be discovered. Another difference is that MTFM is significantly more computationally intensive than the relatively simple Hermes / OTMF algorthms. The repeated metric re-weighting required for real time behaviour detection is therefore an area that requires optimization. We

demonstrated initial, unfiltered Grey Trust assessment using all available metrics (transmitted and received throughput, delay, received signal strength, transmitted power, and packet loss rate), as well as the application of multiple weighting vectors to iteratively emphasise different aspects of trust operation to expose and identify misbehaviour on the network. With significant delays (from seconds to many minutes), in a fading, refractive medium with varying propagation characteristics, the environment is not as predictable or performant as classical MANET TMF deployment environments.

We show that, without significant adaptation, single metric probabilistic estimation based TMFs are ineffective in such an environment. We have shown that existing frameworks are overly optimistic about the nature and stability of the communications channel, and can overlook characteristics that are useful for assessing the behaviour of nodes in the network. This indicates that there is a good case, particularly within constrained MANETs as this, for multi-vector, and even multi-domain trust assessment, where metrics about the communications network and topology would be brought together with information about the physical behaviours and operations of nodes to assess trust.

Also, a significant factor of trust assessment in such a constrained environment, is that there may be long periods where two edge nodes (for instance, $n_0 \rightarrow n_5$) may not interact at all. This can be due to a range of factors beyond malicious behaviour, including simple random scheduling coincidence and intermediate or neighbouring nodes collectively causing long back-off or contention periods. This disconnection hinders trust assessment in two ways; assessing nodes that do not receive timely recommendations may make decisions based on very old data, and malicious nodes have a long dwelling time where they can operate under a reasonable certainty that the TMF will not detect it (especially if the node itself is behaving disruptively). One solution to this would be to move from a stepping-window of trust observations to a continuous trust log, updated on packet reception rather than waiting regular periods for packets to be analysed. Future work will investigate the improvement of weight-based detection algorithms, the stability of GRA under multi-node collusion, the development of real-time outlier detection, and the introduction of physical behavioural metrics into the trust assessment context.

Chapter 7

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

7.1 Introduction

In this chapter we demonstrate the use and operation of a multi-domain trust management framework (MD-TMF) in collaborative marine MANETS 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. We present a methodology for assessing the performance of varying metric sets in detection and differentiation of a range of communications and physical misbehaviours, demonstrating that by utilising information from multiple domains, trust assessment can be more accurate in identifying misbehaviour than in single-domain assessment.

The core part of this chapter was submitted to AAMAS 2016

7.2 Construction of Multi-Domain Trust

A key question in this chapter is to assess the advantages and disadvantages of utilising trust from across domains. This includes a secondary question as to how trust assessments from these domains are most effectively combined.

It is important to clarify what is meant by "effective" in this case; we take the "effectiveness" of any trust assessment framework as consisting of several parts.

1. the *accuracy* of detection and identification of a particular misbehaviour
2. the *timeliness* of such detections

3. the *complexity* of such analysis, including any specific training required
4. the *commonality* of the results of any detections between perspectives (also termed "isomorphism" of results)

7.2.1 Communications Trust Metrics

We use the same trust metrics from [57] that are applicable to the marine environment, i.e. as the simulated modem stack does not operate on the same tiered data-rate approach as used in the 802.11 stack, that metric was not included. Remaining metrics are; Delay, Received and Transmitted power, Received and Transmitted Throughput, and Packet Loss Rate (PLR).

Thus, the metric vector used for communications-trust assessment is;

$$X_{comms} = \{D, P_{RX}, P_{TX}, Tp_{RX}, Tp_{TX}, PLR\} \tag{7.1}$$

7.2.2 Physical Trust Metrics

Three physical metrics are selected to encompass the relative distributions and activities of nodes within the network; Inter-Node Distance Deviation (INDD), Inter-Node Heading Deviation (INHD), and Node Speed. These metrics encapsulate the relative distributions of position and velocity within the fleet, optimising for the detection of outlying or deviant behaviour within the fleet.

Conceptually, INDD is a measure of the average spacing of an observed node with respect to its neighbours. INHD is a similar approach with respect to node orientation.

$$INDD_{i,j} = \frac{|P_j - \sum_x \frac{P_x}{N}|}{\frac{1}{N}\sum_x \sum_y |P_x - P_y| (\forall x \neq y)} \tag{7.2}$$

$$INHD_{i,j} = \hat{v}|v = V_j - \sum_x \frac{V_x}{N} \tag{7.3}$$

$$S_{i,j} = |V_j| \tag{7.4}$$

Thus, the metric vector used for physical-trust assessment is;

$$X_{phy} = \{\text{INDD}, \text{INHD}, S\} \tag{7.5}$$

7.2.3 Metric Weight Analysis Scheme

From (3.16), the final trust values arrived at are dependent on metric values, the weights assigned to each metric, and the structure of the g, b comparison vectors.

This permits the assessment of the significance of different metrics in the detection and identification of different behaviours. For a metric weight vector H, where the metric m_j is emphasised as being twice as important as the other metrics, we form an

initial weighting vector $H' = [h_i...h_M]$ such that $h_i = 1 \forall i \neq j; h_j = 2$. We then scale that vector H' such that $\sum H = 1$ by $H = \frac{H'}{\sum H'}$.

The construction of the g and b vectors from 3.15 depends on the particular metric, e.g. Throughput is positively correlated to trustworthiness and so follows the default construction ($g \mapsto \max, b \mapsto \min$). However, in the case of a metric such as delay, this relationship is inverted, i.e. longer delays indicate less trustworthy activity. In complex environments, the relationship between metrics trustworthiness correlations may not be quite so obvious as the throughput / delay examples. This phenomenon was mentioned by Guo, but was manually configured for each metric for each behaviour and no analytical method for quantatively establishing such relationships has been presented since.

We include both the correlation and relevance of metrics to behaviours by signifying "flipped" metrics (i.e. those with the construction $g \mapsto \min, b \mapsto \max$) by a negative weight.

Using this process we can extract and highlight the primary aspects of an attack by comparing against the deviation from the "fair" result set, i.e. we are interested in the weight schemes that create the largest difference between fair and misbehaving cases.

With the nine selected metrics from across communications and physical behaviours, we can explore this metric space by varying the weights associated with each metric, and choose to emphasise across three levels; i.e. metrics can be ignored or over-emphasised. Naively this results in $3^9 = 19683$ combinations, however as these weights are being normalised, duplicates are introduced, e.g. $[0, 0, 0, 0, 1, 0, 0, 0, 0] \equiv [0, 0, 0, 0, 2, 0, 0, 0, 0]$ leaving 18661 unique weights for analysis.

To assess the performance of a given weight combination (i.e. an optimisation factor), we are initially interested in the metric weight vector that consistently provides the largest deviation in the final trust value T across the cohort, i.e. producing the most clear detection of a node misbehaving in that particular fashion. We approach this as an inverse outlier filtering problem, and select the range outside a $\pm\sigma$ envelope compared to the equivalent weighting in a known "fair" behaviour to assess detection (or comparing to other misbehaviours to assess discrimination). Note that at this point we establish "signatures" of different behaviours rather than optimal detection weights.

We apply a Random Forest regression [56] to assess the relative importance of the selected metrics on relative detectability of malicious behaviour. Random Forest accomplishes this by generating a large number of random regression trees and prune these trees to fit incoming data. A major advantage of Random Forest in this case is that by walking the most successful regression trees, we can acquire an already normalised maximal activation weight for the particular behaviour comparison being tested.

After establishing the importance of weights in particular behaviours, a final weight is arrived at by algorithmically those few metrics that are important, rather than having to further explore the computationally expensive weight-space.

Using this approach we can explore the results of these simulations, condensing the multi-dimensional problem (target / observer / behaviour / metric / time) down to a more tangible level for analysis.

7.3 Results and Discussion

7.3.1 Significance Analysis

First we discuss the results of the Random Forest regression assessment; in Figs 7.1 and 7.2, we show the resultant feature extraction signatures for Comms-only and Physical-only metric selections, and Fig 7.3, these metric spaces are brought together and re-assessed.

It is also interesting to note that in both single-domain cases, there are clear 'signatures' in misbehaviours that don't directly target that domain (P_{RX} in the Physical Shadow and Slowcoach behaviours in Fig 7.1 and $INDD$ in the Selfish Target Selection behaviour in Fig 7.2). This inter-domain activity is to be expected in MANETs in general, where the physical reality of the network (i.e. distance between nodes) directly impacts the behaviour of the logical communications network (i.e. delay between nodes), and is as we will see a useful characteristic for differentiating potential misbehaviours.

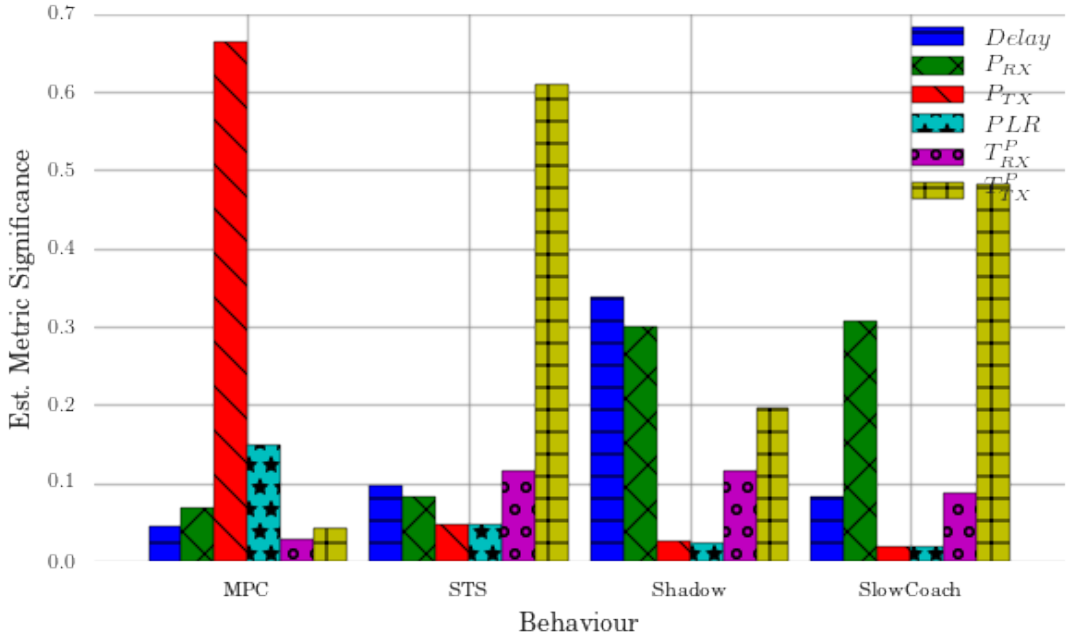


FIGURE 7.1: Plot of X_{comms} Metric Feature Extraction

7.3.2 Weight Assessment

From this significance information we can infer a signature for each behaviour, that can be fed back into the assessment framework, with the aim being to minimise the number

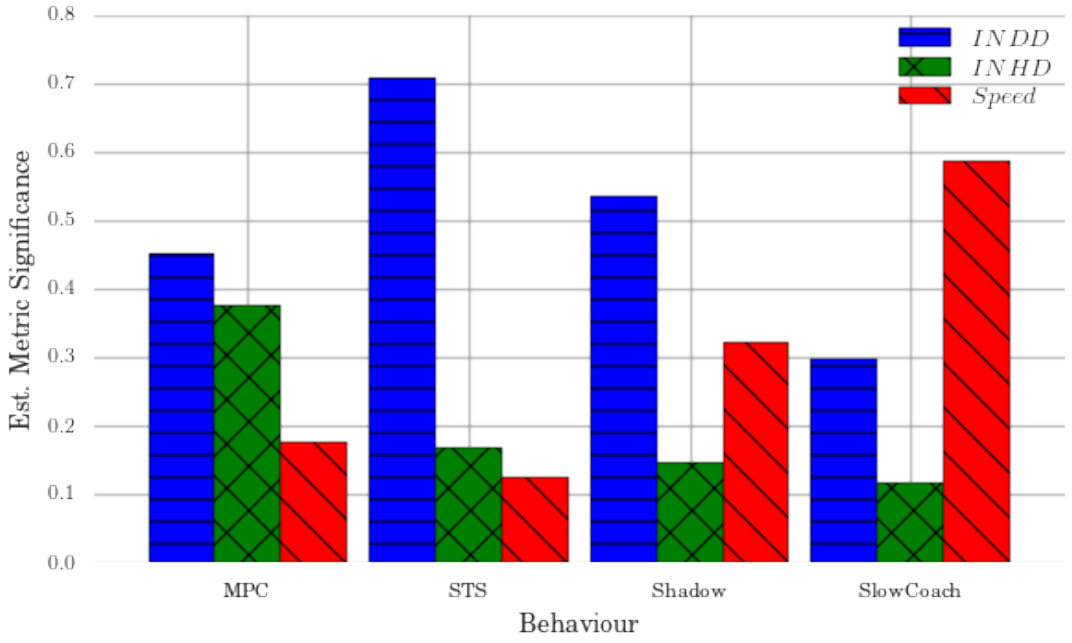


FIGURE 7.2: Plot of X_{phys} Metric Feature Extraction

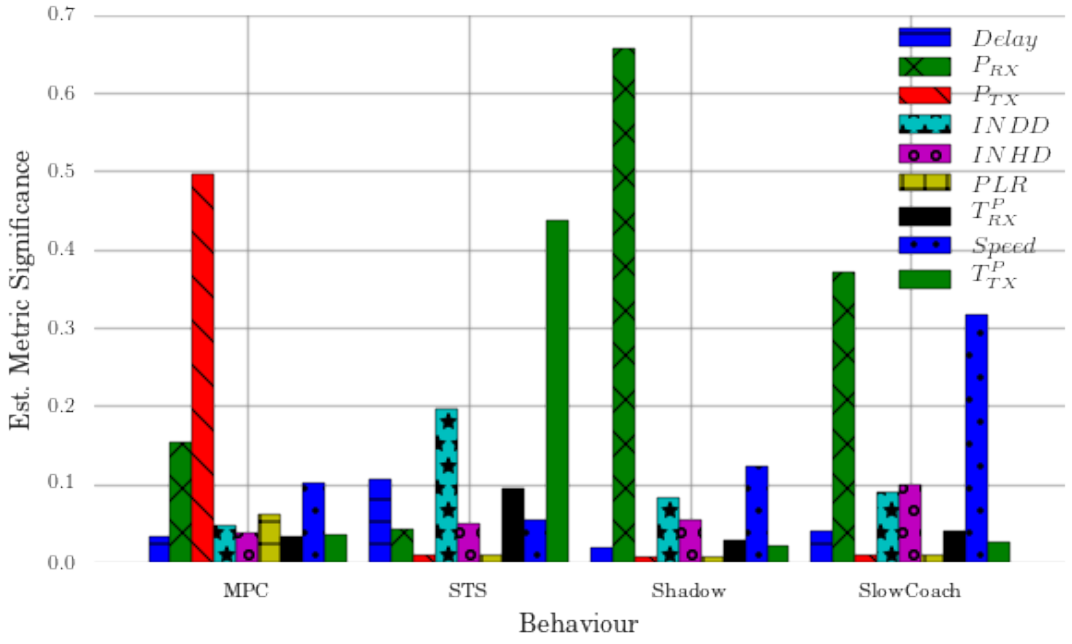


FIGURE 7.3: Multi Domain Relevance assessment of Metric Features

of weight permutations required to come to a conclusion about the behaviour under observation.

We take the feature significances as presented from the regression as baseline weight vectors, however, we have no algorithmically derived approach to the structure of the g,b comparison vectors from (3.16).

TABLE 7.1: ΔT across domains and detected behaviours

Behaviour	MPC	STS	Shadow	SlowCoach	Avg.
Domain					
Full	0.905	0.101	0.499	0.627	0.533
Comms	0.954	0.166	0.287	0.268	0.419
Phys	0.022	0.020	0.421	0.756	0.305
Avg.	0.627	0.096	0.402	0.550	0.419

One option would be to go back to the regression point and expand the combination options to include negative values, however this is combinatorically explosive. Instead, the "significance" weight is permuted against it's possible combinations of "flips", i.e. for $X_s = [0.3, 0.4, 0.01, 0.02, 0.27]$ could also be $X_s^p = [0.3, -0.4, 0.01, 0.02, 0.27]$ and so on. This sign permutation is filtered based on a threshold value (0.01), so for all indices below that threshold will not be permuted on, halving the number of combinations required for each indices eliminated.

The best of these permutations is selected to both maximise the (correct) deviation between each nodes trust perspectives and to minimise the trust value reported for the misbehaving nodes; ΔT max

These weights are applied to untrained data to derive the following results.

An exemplar subset of the results is shows in Figs 7.4-7.9, with the "misbehaving node" highlighted with heavier lines, with any observations about the rest of the cohort faded and dashed. For each node assessment, the mean for that assessment over that time period is also included as a solid / dashed line respectively for clarity.

Comparing Figs 7.4 and 7.5, while there is a reasonable dip in the misbehavor's trust assment, the variance across the cohort is such that this "mistrust" triggering is neither consistent or obvious. Unfortunately this is the case across the STS responses, where in Table 7.1 where we have summarized out general results, STS has by far and away the lowest average ΔT in all domains. Interestingly however is the observation that Comms-only trust performs slightly better than Full trust weighting.

Referring to Figs 7.1 and 7.3, it's clear that the transmitted throughput (T_{TX}^P) is the almost singular feature of this behaviour, due to it's almost completly logical behaviour that is only loosely coupled to the state of the environment. The massive emphasis placed on throughput could only be diminished by putting it together in a larger ensemble.

The other "Primary Communications" behaviour, MPC, is not shown for brevity, but scores comfortably in the 90th percentile range in both full and comms trust assessments.

In Figs 7.6 and 7.7, the misbehaving node is much more obvious than in STS, which is moderatly surprising for a physically-focused behaviour. Further, there is a roughly 20% improvement when incorporating the full metric space.

From Table 7.1, the Shadow behavior is the most consistently detectible behaviour across domains.

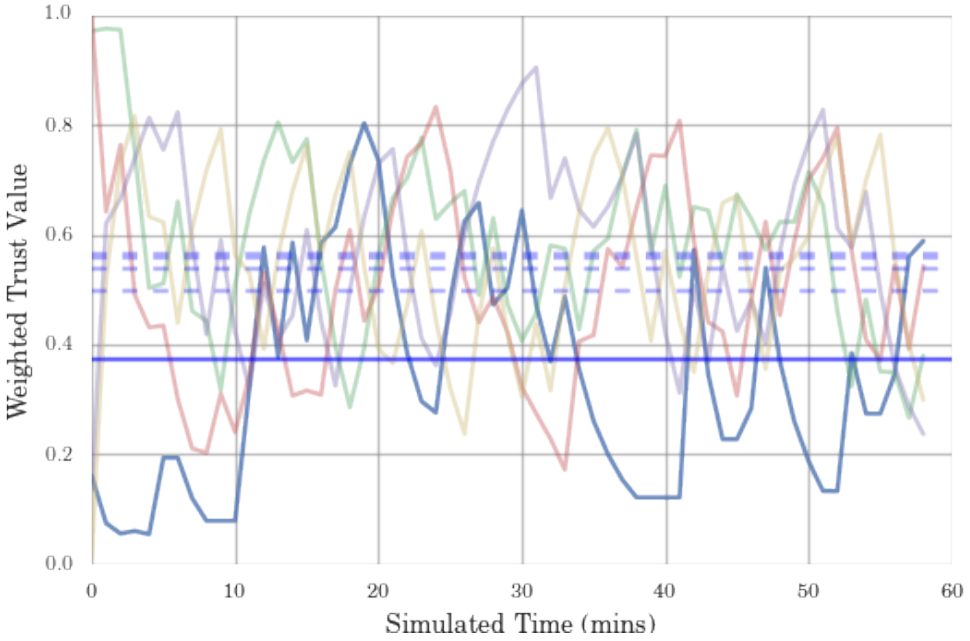


FIGURE 7.4: Selfish(STS) Targeting Comms Metric Trust

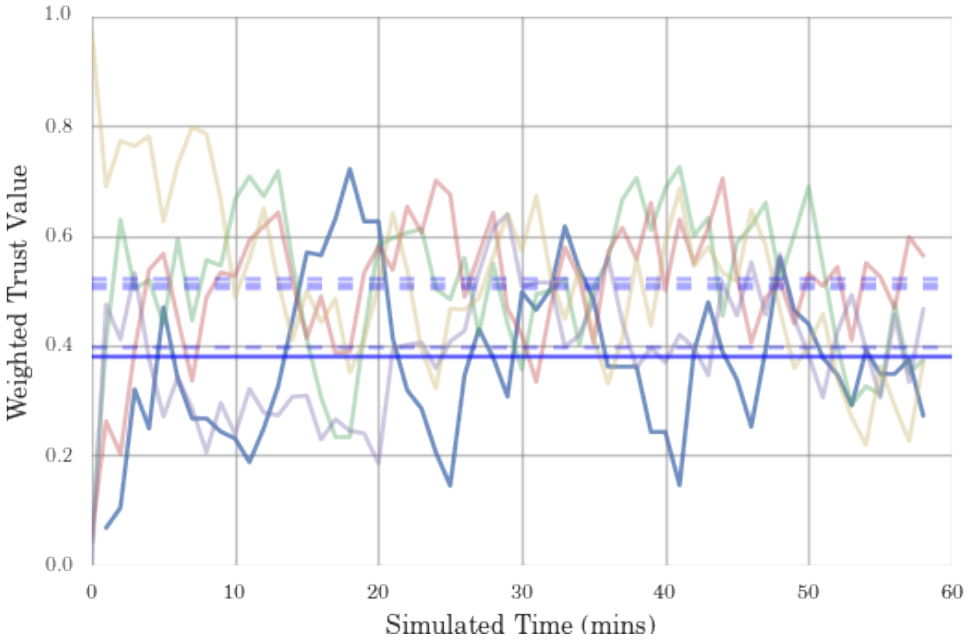


FIGURE 7.5: Selfish(STS) Targeting Full Metric Trust

7.4 Conclusion

In this paper we demonstrate that in harsh environments, multi-domain trust assessment can perform better on average than single-domain counterparts, both in terms of robustness and sensitivity, but also covering a wider region of the potential behaviour space,

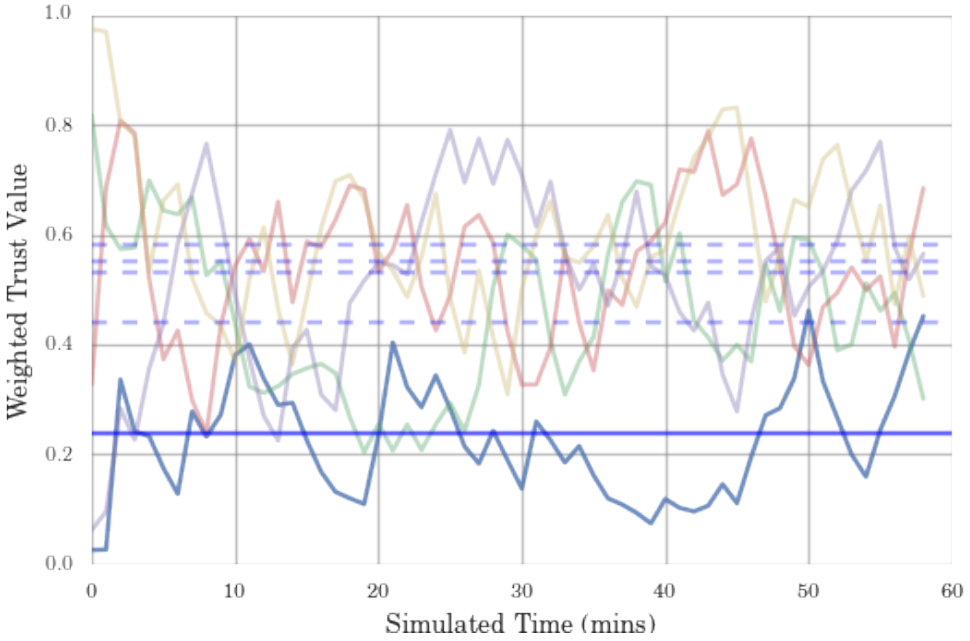


FIGURE 7.6: Shadow Comms Metric Trust

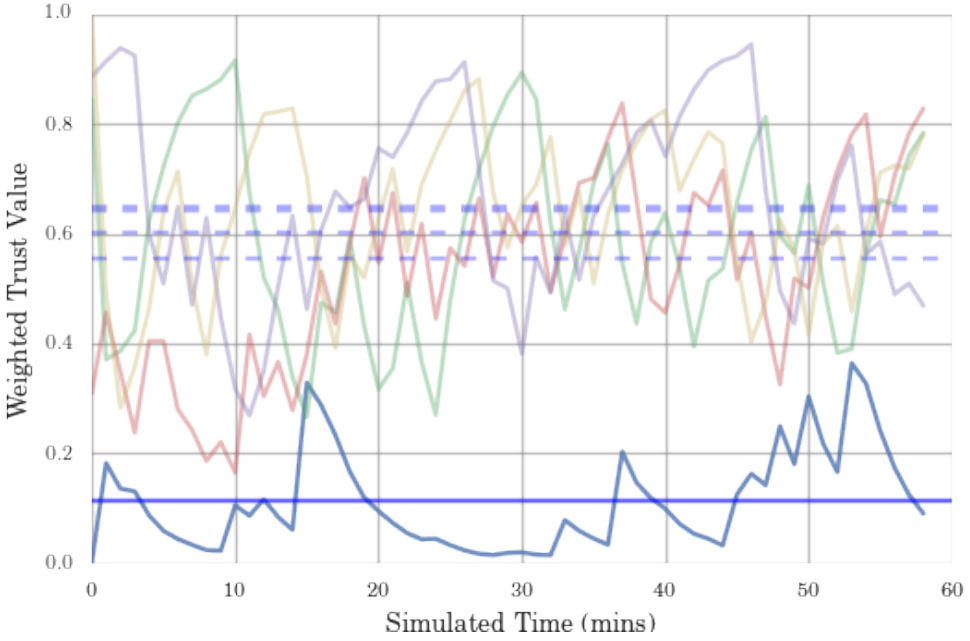


FIGURE 7.7: Shadow Full Metric Trust

The extension of the methodologies of multi-vector trust into the marine space are already demonstrated, however including information from physical observations of actors in a network enables the detection and identification of a much wider range of behaviours. We also demonstrate a method for assessing trust metrics in harsh environments in terms of their relative significance, and a method for establishing classification signatures for misbehaviours.

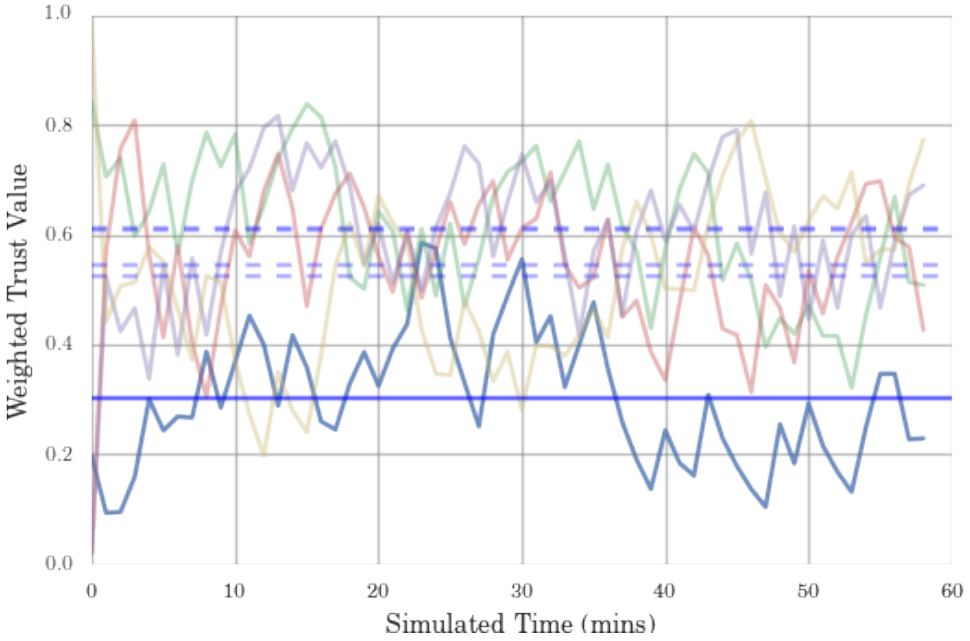


FIGURE 7.8: SlowCoach Comms Metric Trust

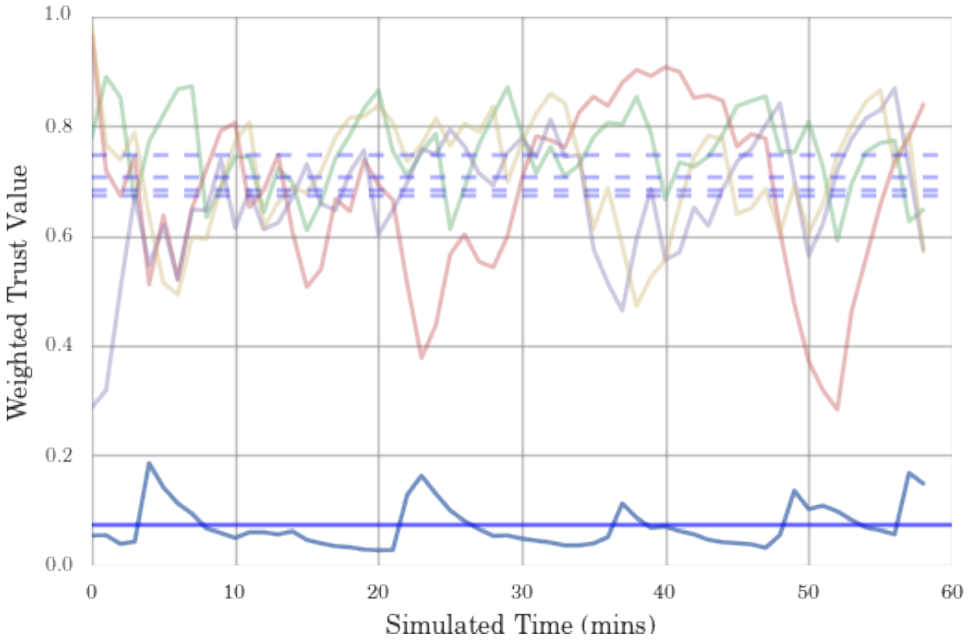


FIGURE 7.9: SlowCoach Full Metric Trust

It is to be noted that this presented method is significantly more computationally intensive than the relatively simple Hermes / OTMF algorithms communications only algorithms, and is exponential in complexity as metrics and/or domains are added. The repeated metric re-weighting required for real time behaviour detection is therefore an area that requires optimization. More work needs to be done to characterise how

worthwhile this approach is compared to a separate synthesis approach where by MTFM-style trust is generated and assessed on a per-domain basis and subsequently fuzed.

For greater fidelity and more optimal results, a wider range of weights can be used in the initial regression step; however this is computationally expensive given that weighting is applied to each perspective (i.e. observer/target node pair) for each trust assessment time step, presenting 15 perspectives at each time interval in the 6 node case.

Every effort has been made to avoid over-training the dataset, using cross validating sampling for regression and "best weight" generation, however more meta-analysis is required to further demonstrate the functionality of this process.

Appendix A

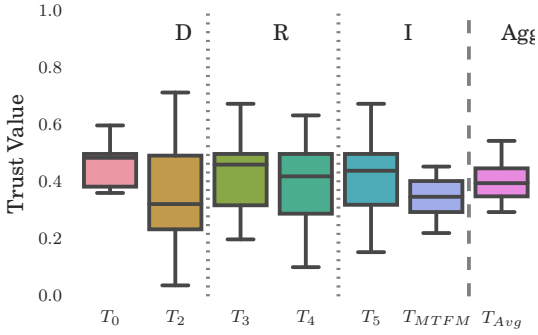
Orphan Sections

A.1 Metric Weighting

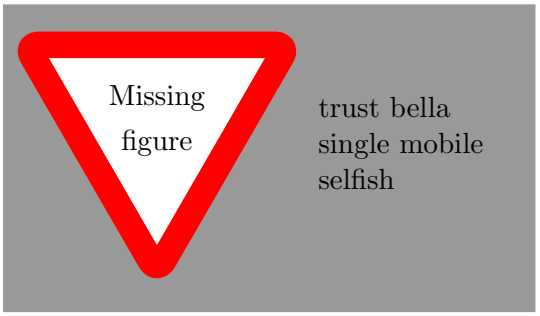
A.2 UNEDITED PROSE: Real Time Grey Systems

Incoming Train of Consciousness

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 [51], we demonstrated a periodic

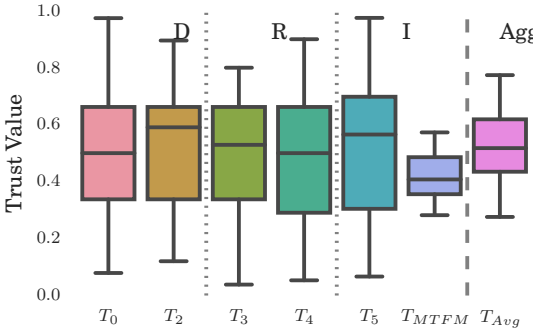


(A) All Nodes Static



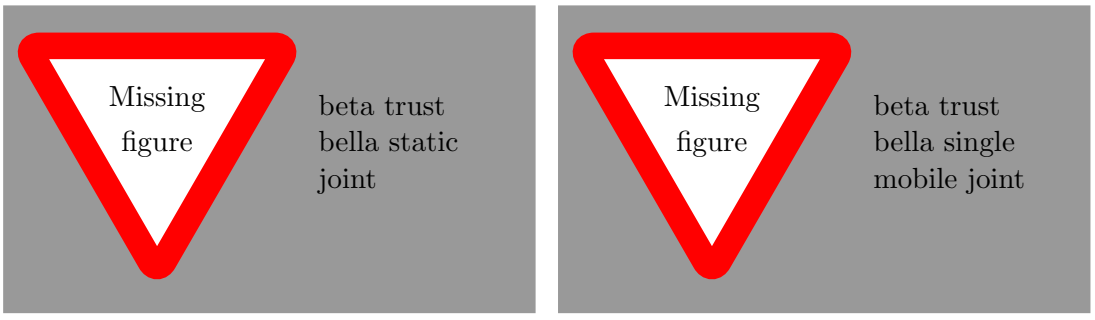
(B) n_1 Randomly Walking

(C) All Nodes but n_1 Randomly Walking



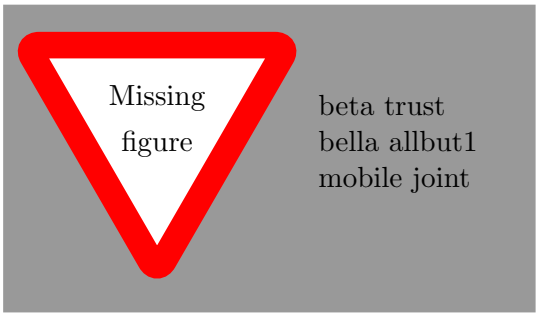
(D) All Nodes Randomly Walking

FIGURE A.1: MTFM Trust assessments for varying mobility options in the selfish case

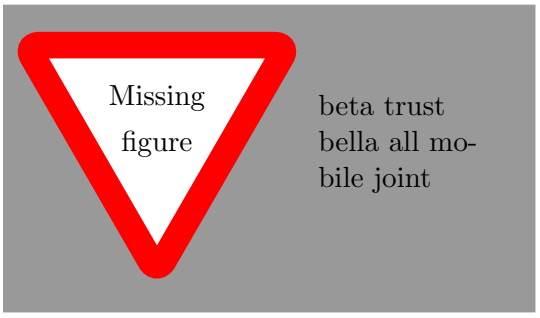


(A) All Nodes Static

(B) n_1 Randomly Walking



(C) All Nodes but n_1 Randomly Walking



(D) All Nodes Randomly Walking

FIGURE A.2: Beta Trust time varying assessments for of $n1$ varying mobility options

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-propagator would be a total mess as I'd imagined that significant-model-breaking would potentially 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.

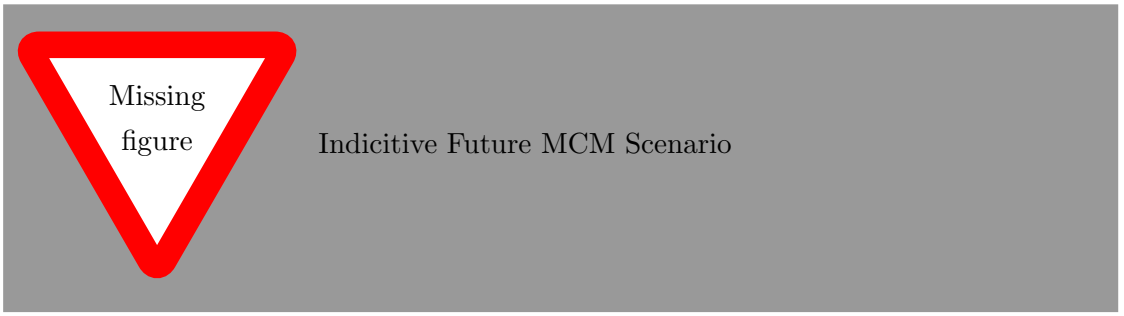
This back propagation will probably be a pain to implement as it'd require a retroactive reassessment of trust and could get really messy if it was interrupt driven, but it's better not to prematurely optimise.

A.3 From end of Defense Trust Conclusions

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)



Indicitive Future MCM Scenario

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.

Bibliography

[1] Jonny Milliken and David Linton. Prioritisation of citizen-centric information for disaster response. *Disasters*, pages n/a—-n/a, 2015. ISSN 1467-7717. doi: 10.1111/disa.12168. URL `http://dx.doi.org/10.1111/disa.12168`.

[2] S Bhargavi and Vishnu Prasad Goranthala. The Impact of Collusion Attacks in WSN with Secure Data Aggregation System. 2(08):213–217, 2015.

[3] J. Jubin and J.D. Tornow. The DARPA packet radio network protocols. *Proc. IEEE*, 75(1):21–32, 1987. ISSN 0018-9219. doi: 10.1109/PROC.1987.13702.

[4] I Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Comput. Networks*, 38(4):393–422, 2002. ISSN 13891286. doi: 10.1016/S1389-1286(01)00302-4. URL `http://linkinghub.elsevier.com/retrieve/pii/S1389128601003024`.

[5] S Corson and J Macker. Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations. RFC 2501, RFC Editor, jan 1999.

[6] Huaizhi Li and Mukesh Singhal. Trust Management in Distributed Systems. *Computer (Long. Beach. Calif).*, 40(2):45–53, 2007. ISSN 00189162. doi: 10.1109/MC.2007.76. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4085622`.

[7] Sonja Buchegger and Jean-Yves Le Boudec. Performance analysis of the CONFIDANT protocol. In *Proc. 3rd ACM Int. Symp. Mob. ad hoc Netw. Comput. - MobiHoc '02*, pages 226–236. ACM Press, 2002. ISBN 1581135017. doi: 10.1145/513800.513828. URL `http://dl.acm.org/citation.cfm?id=513800.513828`.

[8] John D Lee and Katrina A See. Trust in automation: designing for appropriate reliance. *Hum. Factors*, 46(1):50–80, 2004. ISSN 0018-7208. doi: 10.1518/hfes.46.1.50.30392.

[9] Tetsushi Okumura, Jeanne M. Brett, William W. Maddux, and Peter H. Kim. Cultural Differences in the Function and Meaning of Apologies. *Int. Negot.*, 16:405–425, 2011. ISSN 1382-340X. doi: 10.1163/157180611X592932.

[10] Roger C Mayer, James H Davis, and F David Schoorman. An Integrative Model of Organizational Trust. *Acad. Manag. Rev.*, 20(3):709–734, jul 1995. ISSN 03637425. doi: 10.2307/258792. URL http://www.jstor.org/stable/258792.

[11] Julian B Rotter. A new scale for the measurement of interpersonal trust1. *J. Pers.*, 35(4):651–665, 1967. ISSN 1467-6494. doi: 10.1111/j.1467-6494.1967.tb01454.x. URL http://dx.doi.org/10.1111/j.1467-6494.1967.tb01454.x.

[12] Yan Lindsay Sun, Rhode Island, Z Han, and K J R Liu. Defense of trust management vulnerabilities in distributed networks. *IEEE Commun. Mag.*, 46(2):112–119, 2008. URL http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=4473092.

[13] R. Alami, R. Chatila, S. Fleury, M. Ghallab, and F. Ingrand. An Architecture for Autonomy. *Int. J. Rob. Res.*, 17:315–337, 1998. ISSN 0278-3649. doi: 10.1177/027836499801700402.

[14] George A Bekey. *Autonomous robots : from biological inspiration to implementation and control*. 2005. ISBN 0262025787. URL http://books.google.com/books?hl=en{&}amp;lr={&}amp;id=3xwfia2DpmoC{&}amp;oi=fnd{&}amp;pg=PR13{&}amp;dq=Autonomous+Robots+From+Biological+Inspiration+to+Implementation+and+Control{&}amp;ots=WxngXPbihr{&}amp;sig=7G8VA4GRaUOwc0sAFbfPi5uAK18.

[15] Stan Franklin and Art Graesser. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In *Intell. agents III agent Theor. Archit. Lang.*, pages 21–35. Springer, 1997.

[16] H. M. Huang. Autonomy Levels for Unmanned Systems (ALFUS) Framework Volume I : Terminology Unmanned Systems Working Group Participants 1 National Institute of Standards and Technology. *Framework*, I(September):29, 2004.

[17] R.R. Murphy. *Introduction to AI robotics*, volume 108. 2000. ISBN 0262133830. doi: 10.1111/j.1464-410X.2011.10513.x. URL http://www.ncbi.nlm.nih.gov/pubmed/21917105.

[18] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach, 3rd edition*. 2009. ISBN 0136042597. doi: 10.1017/S0269888900007724. URL http://portal.acm.org/citation.cfm?id=1671238{&}coll=DL{&}dl=GUIDE{&}CFID=190864501{&}CFTOKEN=29051579$\delimiter"026E30F$npapers2://publication/uuid/4B787E16-89F6-4FF7-A5E5-E59F3CFEFE88.

[19] Sebastian Thrun. Toward a Framework for Human-Robot Interaction. *Human-Computer Interact.*, 19:9–24, 2004. ISSN 0737-0024. doi: 10.1207/s15327051hci1901{\&}2{_}2.

[20] Michael Wooldridge and Nicholas R. Jennings. Intelligent agents: theory and practice, 1995. ISSN 0269-8889.

[21] Thomas B. Sheridan and William L. Verplank. Human and Computer Control of Undersea Teleoperators. *ManMachine Syst. Lab Dep. Mech. Eng. MIT Grant N0001477C0256*, page 343, 1978.

[22] M R Endsley and D B Kaber. *Level of automation effects on performance, situation awareness and workload in a dynamic control task.*, volume 42. 1999. ISBN 0014013991. doi: 10.1080/001401399185595.

[23] NATO Standardization Office. STANAG 4586 STANDARD INTERFACES OF UAV CONTROL SYSTEM (UCS) FOR NATO UAV INTEROPERABILITY Ed: 3. Technical report, NATO, Brussels, Belgium, 2012. URL `http://nso.nato.int/nso/zPublic/stanags/current/4586eed03.pdf`.

[24] Mary L. Cummings, Sylvain Bruni, and Paul J. Mitchell. Chapter 2¡BR¿ Human Supervisory Control Challenges in Network-Centric Operations, 2010. ISSN 1557234X.

[25] Jenay M Beer, Arthur D Fisk, and Wendy a Rogers. Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction. *J. Human-Robot Interact.*, 3(2): 74–99, 2014. ISSN 2163-0364. doi: 10.5898/JHRI.3.2.Beer.

[26] Neya Systems LLC. The JAUS Toolset. URL `http://jaustoolset.org/`.

[27] American Society of Testing and Materials. ASTM F2500 - 07 Standard Practice for Unmanned Aircraft System (UAS) Visual Range Flight Operations. Technical report, 2007. URL `http://www.astm.org/Standards/F2500.htm`.

[28] American Society of Testing and Materials. ASTM F2541-06 Standard Guide for Unmanned Undersea Vehicles (UUV) Autonomy and Control. Technical report, 2006. URL `http://www.astm.org/Standards/F2541.htm`.

[29] Nick Johnson, Pedro Patron, and David Lane. The importance of trust between operator and AUV: Crossing the human/computer language barrier. *Ocean. 2007 - Eur.*, pages 1–6, jun 2007. doi: 10.1109/OCEANSE.2007.4302408. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4302408`.

[30] Nicholas A. R. Johnson and David M. Lane. Narrative monologue as a first step towards advanced mission debrief for AUV operator situational awareness. *2011 15th Int. Conf. Adv. Robot.*, pages 241–246, 2011. doi: 10.1109/ICAR.2011.6088618.

[31] Jessie Y. C. Chen, Michael J. Barnes, and Michelle Harper-Sciarini. Supervisory Control of Multiple Robots: Human-Performance Issues and User-Interface Design. *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, 41(4):435–454, 2011. ISSN 1094-6977.

[32] Aaron Mehta. Political, Financial Threads Underscore German Euro Hawk Saga. *Def. News*, jun 2013.

[33] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The Eigentrust algorithm for reputation management in P2P networks. *12th Int. Conf. World Wide Web (WWW)*, page 640, 2003. ISSN 1581136803. doi: 10.1145/775240.775242. URL `http://portal.acm.org/citation.cfm?doid=775152.775242`.

[34] Charikleia Zouridaki, Brian L Mark, Marek Hejmo, and Roshan K Thomas. A quantitative trust establishment framework for reliable data packet delivery in MANETs. *Proc. 3rd ACM Work. Secur. ad hoc Sens. networks*, pages 1–10, 2005. ISSN 0926227X. doi: 10.1145/1102219.1102222.

[35] 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 (4):108–114, apr 2007. ISSN 01636804. doi: 10.1109/MCOM.2008.4481349. URL `http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=4212452http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4481349`.

[36] Jin-hee Cho, Ananthram Swami, and Ing-ray Chen. A survey on trust management for mobile ad hoc networks. *Commun. Surv. & Tutorials*, 13(4):562–583, 2011. URL `http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5604602`.

[37] MEG E G Moe, BE E Helvik, and SJ J Knapskog. TSR: Trust-based secure MANET routing using HMMs. *. . . symposium QoS Secur. . . .*, pages 83–90, 2008. URL `http://dl.acm.org/citation.cfm?id=1454602`.

[38] 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*, pages 305–311, 2008. doi: 10.1109/LCN.2008.4664184. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4664184`.

[39] 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*, pages 142–149, 2011. doi: 10.1109/TrustCom.2011.21. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6120813`.

[40] Fengchao Zuo. Determining Method for Grey Relational Distinguished Coefficient. *SIGICE Bull.*, 20(3):22–28, jan 1995. ISSN 0893-2875. doi: 10.1145/202081.202086. URL `http://doi.acm.org/10.1145/202081.202086`.

[41] Liang Hong Liang Hong, Wu Chen Wu Chen, Li Gao Li Gao, Guoqing Zhang Guoqing Zhang, and Cai Fu Cai Fu. Grey theory based reputation system for secure

neighbor discovery in wireless ad hoc networks. *Futur. Comput. Commun. (ICFCC), 2010 2nd Int. Conf.*, 2, 2010. doi: 10.1109/ICFCC.2010.5497609.

[42] Andrea Caiti. Cooperative distributed behaviours of an AUV network for asset protection with communication constraints. *Ocean. 2011 IEEE-Spain*, 2011. URL `http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6003463`.

[43] Surya Pavan, Kumar Gudla, and N Preeti. An Overview of Reputation and Trust in Multi Agent System in Disparate Environments. 5(3):498–504, 2015.

[44] 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(4):17, 2006. ISSN 15591662. doi: 10.1145/1161039.1161045. URL `http://portal.acm.org/citation.cfm?doid=1161039.1161045`.

[45] Milica Stojanovic. On the relationship between capacity and distance in an underwater acoustic communication channel, 2007. ISSN 15591662. URL `http://www.mit.edu/{~}millitsa/resources/pdfs/bwdx.pdf`.

[46] Xavier Lurton. *An introduction to underwater acoustics: principles and applications*. Springer Praxis Books. Springer Berlin Heidelberg, 2002. ISBN 9783540784807. URL `https://books.google.fr/books?id=PFXgLQAACAAJ`.

[47] Kenneth V. Mackenzie. Nineterm equation for sound speed in the oceans. *J. Acoust. Soc. Am.*, 70(3):807, 1981. ISSN 00014966. doi: 10.1121/1.386920.

[48] R F W Coates. *Underwater acoustic systems*. A Halstead Press book. John Wiley & Sons Canada, Limited, 1989. ISBN 9780470215449. URL `https://books.google.co.uk/books?id=0qUeAQAAIAAJ`.

[49] Chiara Petrioli and Roberto Petroccia. SUNSET: Simulation, emulation and real-life testing of underwater wireless sensor networks. *Proc. IEEE UComms 2012*, 2012. URL `http://reti.dsi.uniroma1.it/UWSN{_}Group/publications/pdf/2012/sunset.pdf`.

[50] Sifeng Liu and Yi Lin. *Grey System Theory and Application*. Number 1. Springer-Verlag Berlin Heidelberg, 2011. ISBN 978-1-61284-490-9. doi: 10.1109/GSIS.2011.6044018. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6044018$\delimiter"026E30F$nhttp://www.springer.com/physics/complexity/book/978-3-642-16157-5`.

[51] Andrew Bolster and Alan Marshall. Single and Multi-Metric Trust Management Frameworks for use in Underwater Autonomous Networks. In *Trust. 2015*, 2015.

[52] Klaus Müller and Tony Vignaux. SimPy: Simulating Systems in Python. *ONLamp.com Python DevCenter*, feb 2003. URL `http://www.onlamp.com/pub/a/python/2003/02/27/simpy.html?page=2`.

[53] Josep Miquel and Jornet Montana. AUVNetSim: A Simulator for Underwater Acoustic Networks. *Program*, pages 1–13, 2008. URL `http://users.ece.gatech.edu/jmjm3/publications/auvnetsim.pdf`.

[54] Andrej Stefanov and Milica Stojanovic. Design and performance analysis of underwater acoustic networks. *IEEE J. Sel. Areas Commun.*, 29(10):2012–2021, 2011. ISSN 07338716. doi: 10.1109/JSAC.2011.111211.

[55] Kaixin Xu, Mario Gerla, Sang Bae, and Hoc Networks. Effectiveness of RTS / CTS Handshake in IEEE. *... , 2002. Globecom'02. Ieee*, 56:1–14, 2002. ISSN 15708705. doi: 10.1049/el. URL `http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=1188044`.

[56] L Breiman. Random forests. *Mach. Learn.*, pages 5–32, 2001. ISSN 0885-6125. doi: 10.1023/A:1010933404324. URL `http://link.springer.com/article/10.1023/A:1010933404324`.

[57] Ji Guo. Trust and Misbehaviour Detection Strategies for Mobile Ad hoc Networks. 2012.