A Multi-Domain Trust Framework for Autonomous Mobile Networks in Harsh Environments

Registration ID#176

ABSTRACT

Trust Management Frameworks (TMFs) are being used to improve the efficiency, security, and reliability of decentralized and distributed autonomous systems. Techniques have been developed for high-speed, uncontended environments such as terrestrial 802.11 MANETs. However, these do not perform well in sparse / harsh environments such as those found in Underwater Acoustic Networks (UANs), where network nodes experience significant and variable delays, comparatively low data rates, large contention periods, and considerable multi-path artefacts.[1]

In such sparse networks, trust establishment based on statistical observations of success/failure events become unstable and ineffective in detecting or identifying misbehaviours. Additionally, these methodologies focus solely on the communications activities of entities and do not incorporate information from other domains, such as physical mobility.

In this paper we demonstrate the use and operation of a multi-domain trust management framework (MD-TMF) using UANs as an exemplar application. We present a preliminary 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.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms Management Performance Reliability Security

Keywords

MANET, Underwater, Simulation, Trust

1. INTRODUCTION

With the increasing integration of autonomy into modern cyber-physical systems, Trust Management Frameworks (TMFs) are being applied to safeguard the efficiency, security, and reliability of decentralised and distributed networks of autonomous systems, from highway-bound autonomous vehicles to aerial battlefield drones. Classical implementations of trust management in Mobile Ad-Hoc Networks (MANETs) have focused solely on observations from the communications domain to make trust assessments. However, these methods are not as effective in applications exhibiting sparse, delayed, or otherwise challenged communications environments[18].

In the majority of Trusted MANET implementations, a free space RF communications protocol such as 802.11 is used. By their nature, such implementations rely on relatively high bandwidth, low noise, low latency, high channel occupancy where contention is tolerable. In contrast; in underwater environments, communications is sparse, delayful, noisy, and very prone to destructive contention. 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, cheaper, almost disposable autonomous underwater vehicles (AUVs), particularly in defense, ecological and petrochemical fields. . These increasingly decentralised applications present unique threats against trust management.

Trust Management Frameworks (TMFs) provide information to assist the estimation of future states and actions of nodes within networks. This information is used to optimize the performance of a network against malicious, selfish, or defective misbehaviour by one or more nodes. Previous research has established the advantages of implementing TMFs in 802.11 based MANETs, particularly in terms of preventing selfish operation in collaborative systems [9], and maintaining throughput in the presence of malicious actors [3]. Most current TMFs use a single type of observed action to derive trust values, typically successfully delivered or forwarded packets. These observations then inform future decisions of individual nodes, for example, route selection [10].

Recent work has demonstrated the use of a number of metrics to form a "vector" of trust. The Multi-parameter Trust Framework for MANETs (MTFM) [6], uses a range of communications metrics beyond packet delivery/loss rate (PLR) to assess trust. This vectorized trust also allows a system to detect and identify the tactics being used to undermine

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), John Thangarajah, Karl Tuyls, Stacy Marsella, Catholijn Jonker (eds.), May 9–13, 2016, Singapore.*

or subvert trust. This method as been previously applied to the marine space, comparing against a selection of existing communications TMFs [1] showing that MTFM is more effective at detecting misbehaviours in sparse environments. This paper builds upon that work to encompass physical as well as communications observations in the establishment of trust and the detection and classification of misbehaviours across both physical and communications domains. An example area of application is the underwater marine environment, where extreme challenges to communications are present (propagation delays, frequency dependent attenuation, fast and slow fading, refractive multi-path distortion, etc.).

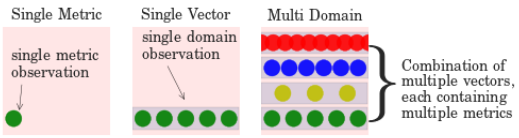

Figure 1: The inclusion of additional metrics and domains in trust assessment reduces the systems exposed threat surface

This paper is laid out as follows; in Section 2 we discuss Trust and TMFs, defining out terminology and reviewing the justifications for the use and development of TMFs in harsh environments such as UANs. In Section 3 we review selected features of the underwater communications channel, highlighting particular challenges against terrestrial equivalents. In Section 4 we review the findings of [1] and establish experimental parameters and simulated behaviours under assessment. In Section 5 we present our analysis method for assessing misbehaviour using MTFM, and intermediate results of the independent detection of physical and communications misbehaviours using single-domain observations. In Section 6 we demonstrate results from multi-domain MTFM and discuss the significance of these findings in terms of detection and classification of cross-domain misbehaviour sets.

2. TRUST MANAGEMENT FRAMEWORKS

2.1 Trust in Connected Systems

For a term that is so common in every-day speech, Trust is a challenging discussion area, particularly given the wealth of proposed definitions [8]. Beyond these 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.

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

In the context of trust between and among collectives of autonomous systems, it is the quantifiable assessment of the expected reliability of an entity to perform a given task on request by another entity. This trust assessment can be targeted towards the reliability of a single task or as a general indicator of "trustworthiness" of a range of activities.

In many systems, this "trust" is analogous to "security", and these terms are often conflated, however in this case we are not necessarily concerned with the security of a system so much as it's runtime reliability in operation.

Within networked systems, the aspect of operation that usually need to be protected from malicious operation or failure states most is bandwidth and congestion, i.e. optimising the network to maximum energy/time efficiency for transiting information. In this respect, the Router Information Protocols (RIP) can be considered very basic Trust Management Frameworks, such that they optimise routing structures within a network based on (binary) link availability.[12].

However, in more complex systems and application, communications optimality is not necessarily the only aspect of operation under threat. Particularly in Mobile Ad-Hoc Networks, individual node constraints such as energy supply and any other consumables present another exploitable vector for attack. Further, in applications such as persistent survey and observation, operational considerations such as the end-to-end delay in reporting or the reduction in the timely observed area may be threatened by misbehaviour.

2.2 Trust in Conventional MANETs

The distributed and dynamic nature of MANETs mean that it is difficult to maintain a trusted third party (TTP) systems or evidence based trust system such as Certificate Authorities (CA) or Public Key Infrastructure (PKI). Distributed trust management frameworks aim to detect, identify, and mitigate the impacts of malicious actors by distributing per-node assessments and opinions to collectively police behaviour.

Various models and algorithms for describing trust and developing trust management in distributed systems, P2P communities or wireless networks have been considered. *Hermes Trust Establishment Framework* uses a Bayesian Beta function to model per-link Packet Loss Rate (PLR) over time, combining "Trust" and "Confidence of Assessment" into a single value [21]. *Objective Trust Management Framework* (OTMF) builds upon Hermes and distributes node observations across the network [10], however does not appropriately combat multi-node-collusion in the network [4]. *Trust-based Secure Routing* demonstrated an extension to Dynamic Source Routing (DSR), incorporating a Hidden Markov Model of sub-networks, reducing the efficacy of Byzantine attacks such as black-hole routing [15]. *CONFIDANT* presented an approach using a probabilistic estimation of PLR, similar to OTMF, also introducing a topology aware weighting scheme and also weighting trust assessments based on historical experience of the reporter [3]. Fuzzy Trust-Based Filtering uses Fuzzy Inference to adapt to malicious recommenders using conditional similarity to classify performance with overlapping fuzzy set membership, filtering assessments across a network [11]. *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[6].

These TMFs (excluding MTFM) can be generalised as single-value estimation based on a binary input state (suc-

cess or failure of packet delivery) and generating a probabilistic estimation of the future states of that input (how likely is it that the next transmission will reach it's destination?). 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.

As an example; The Hermes trust establishment framework [21] uses Bayesian reasoning to generate a posterior distribution function of "belief", or trust, given a sequence of observations of that behaviour, $p(B|O)(1)$.

$$p(B|O) = \frac{p(O|B) \times p(B)}{\rho} \tag{1}$$

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) 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 (3) and these measurements are combined to form a "trustworthiness" value T (4).

$$t_i \to E[\text{beta}(p|\alpha,\beta)] = \frac{\alpha_i}{\alpha_i + \beta_i} \tag{2}$$

$$c_i = 1 - \sqrt{\frac{12\alpha_i\beta_i}{(\alpha_i+\beta_i)^2(\alpha_i+\beta_i+1)}} \tag{3}$$

$$T_i = 1 - \frac{\sqrt{\frac{(t_i-1)^2}{x^2} + \frac{(c_i-1)^2}{y^2}}}{\sqrt{\frac{1}{x^2} + \frac{1}{y^2}}} \tag{4}$$

In (4), x and y are constants, used weight the two-dimensional polar mapping of trust and confidence assessments (t_i, c_i), and from [21], 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.

Multi-Parameter Trust Framework for MANETS (MTFM) extends this single-parameter approach, applying Grey Relational Analysis [22] to provide cohort based normalization of a range of disparate metrics at runtime, providing a "grade" of trust compared to other observed nodes, while maintaining the ability to reduce trust valued down to a stable assessment range for decision support without requiring a-priori environmental or metric characterisation. This presents a stark difference between the previously discussed 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. [6] demonstrated the ability of grey relational analysis (GRA) [22] 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

$$\theta^t_{k,j} = \frac{\min_k |a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|}{|a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|}$$
$$\phi^t_{k,j} = \frac{\min_k |a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|}{|a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|} \tag{5}$$

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 $\{a^t_{k,j}k = 1, 2 \dots K\}$, normally $g_j = \max_k(a^t_{k,j})$, $b_j = \min_k(a^t_{k,j})$ where each metric is monotonically positive for trust assessment.

Weighting can be applied before generating a scalar value (6) allowing the detection and classification of misbehaviours.

$$[\theta^t_k, \phi^t_k] = \left[\sum_{j=0}^{M} h_j\theta^t_{k,j}, \sum_{j=0}^{M} h_j\phi^t_{k,j}\right] \tag{6}$$

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 (5) the $[\theta, \phi]$ values are reduced into a scalar trust value by $T^t_k = (1 + (\phi^t_k)^2/(\theta^t_k)^2)^{-1}$ [7]. MTFM combines this GRA with a topology-aware weighting scheme (7) and a fuzzy whitenization model (8).

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. 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.

$$\begin{aligned} T^{MTFM}_{i,j} =& \frac{1}{2} \cdot \max_s\{f_s(T_{i,j})\}T_{i,j} \\ &+ \frac{1}{2}\frac{2|N_R|}{2|N_R| + |N_I|} \sum_{n \in N_R} \max_s\{f_s(T_{i,n})\}T_{i,n} \\ &+ \frac{1}{2}\frac{|N_I|}{2|N_R| + |N_I|} \sum_{n \in N_I} \max_s\{f_s(T_{i,n})\}T_{i,n} \end{aligned} \tag{7}$$

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:

$$f_1(x) = -x + 1$$
$$f_2(x) = \begin{cases} 2x & \text{if } x \leq 0.5 \\ -2x + 2 & \text{if } x > 0.5 \end{cases} \quad (8)$$
$$f_3(x) = x$$

In the case of the terrestrial communications network used in [6], 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 in terrestrial 802.11 based MANETs, MTFM provided increased variation in trust assessment over time, providing more information about the nodes' behaviours than packet delivery probability alone can. By weighting the metrics used in MTFM it was shown that the trust assessments could be used to identify the style of misbehaviour being performed within the network, and by whom. It has been demonstrated that MTFM is a strong candidate for trust assessment in the harsh marine communications environment [1].

2.3 Multi-Metric Single Domain Trust in Harsh Environments

It has been demonstrated that classical, single metric, MANET Trust Management Frameworks are not directly suitable to the sparse, noisy, and dynamic underwater medium, however MTFM shows promise in it's multi-metric normalisation approach when compared to the performance of OTMF and Hermes in the same scenarios [1]. In a simulated underwater environment, MTFM was able to discriminate between communications behaviours using communications metrics by exploring the metric space by weight variation, allowing the detection and classification of the malicious behaviours. With significant end-to-end 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. Without significant adaptation, single metric probabilistic estimation based TMFs are ineffective in such an environment. This is because 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 for bringing together information from the commmunications and physical domains together to assess and monitor trustworthiess, particularly within constrained MANETs as in the underwater acoustic realm.

3. MARINE ACOUSTIC COMMUNICATIONS

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 its interfaces with the seabed / air. The resultant difficulties include; long propagation delays, 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, multi-path caused by reflective medium interfaces, variations in propagation speed due to depth dependant effects (salinity, temperature, and pressure), and subsequent refractive spreading and lensing due to that same propagation variation [17].

The attenuation that occurs in an underwater acoustic channel over a distance d for a signal about frequency f in linear power is given as $A_{\text{aco}}(d, f) = A_0 d^k a(f)^d$ and in dB form as;

$$10 \log A_{\text{aco}}(d, f)/A_0 = k \cdot 10 \log d + d \cdot 10 \log a(f) \quad (9)$$

where A_0 is a normalising constant, k is a spreading factor (commonly taken as 1.5 [20]), and $a(f)$ is the absorption coefficient, approximated using Thorp's formula [19]

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

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

Comparing $A_{aco}(d, f)$ with the RF Free-Space Path Loss model ($A_{\text{RF}}(d, f) \approx \left(\frac{4\pi df}{c}\right)^2$), the impact of range on signal power is exponential underwater, rather than quadratic in terrestrial RF ($A_{\text{aco}} \propto f^{2d}$ vs $A_{\text{RF}} \propto (df)^2$). While both frequency dependant factors are quadratic, approximating the factors in (10), $f \propto A_{\text{aco}}$ is at least 4 orders of magnitude higher than $f \propto A_{\text{RF}}$

4. ANALYSIS, DESIGN AND PER-DOMAIN RESULTS

4.1 Simulation and Scenario Generation

To investigate the operation of a fully mobile network of six nodes, each kinematically modelled on the commonly used REMUS100 AUV platform [13] in the marine environment, simulations were conducted using a Python based framework, SimPy [16], with a network stack built upon AUVNetSim [14], using transmission parameters (Table 1) taken from and validated against [20] and [19].

Four scenarios were developed to assess both communications and physical domains where one node within the fleet was 'misbehaving' (n_m). In cases where n_m is specifically targeting another node in the fleet, that node is denoted as n_t.

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

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

3. Shadowing, where n_m is not aware of the pre-planned mission paths and is instead simply following it's neighbours.

4. Slow Coach, where n_m is experiencing a simulated power-train failure that reduces it's acceleration and top speed, analogous to a fouled propeller.

The default scenario is also simulated where nodes participate fairly in the network and follow a collaborative survey mobility pattern.

From these simulations, we attain the per-node recorded positions, ($P_i = [x, y, z] \forall t$) as well as each nodes estimations of it's neighbours positions $P_{i,j} = [x, y, z] \forall t$) (which assumes that all nodes are fairly reporting their positions compactly at each transmission), and each nodes trust metric observations of it's neighbours; $A^t_{i,j} = [x_{i,j} \forall x \in X]$ where X are the selected trust metrics.

Table 1: System model constraints

Parameter	Unit	Value
Simulated Duration	s	18000
Initial Separation	m	300
Trust Sampling Period	s	600
Simulated Area	km^2	0.7-4
Transmission Range	km	1.5
Physical Layer	Acoustic	
Propagation Speed	m/s	1490
Center Frequency	Hz	2×10^4
Bandwidth	Hz	1×10^4
MAC Type		CSMA/CA
Routing Protocol		FBR
Max Speed	ms^{-1}	1.5
Max Data Rate	bps	≈ 240
Packet Size	bits	9600
Single Message Duration	s	32
Single Message Size	bits	9600

4.2 Communications Trust Metrics

We use the same trust metrics from [5] 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;

$$X_{comms} = \{D, P_{RX}, P_{TX}, Tp_{RX}, Tp_{TX}, PLR\} \quad (11)$$

4.3 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.

$$INDD_{i,j} = \frac{|P_j - \sum_x \frac{P_x}{N}|}{\frac{1}{N}\sum_x \sum_y |P_x - P_y| (\forall x \neq y)} \quad (12)$$

$$INHD_{i,j} = \hat{v} | v = V_j - \sum_x \frac{V_x}{N} \quad (13)$$

$$S_{i,j} = |V_j| \quad (14)$$

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

$$X_{phy} = \{\text{INDD}, \text{INHD}, S\} \quad (15)$$

4.4 Metric Weight Analysis Scheme

From (6), 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 5 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 [2] 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.

5. RESULTS AND DISCUSSION

5.1 Significance Analysis

First we discuss the results of the Random Forest regression assessment; in Figs 2 and 3, we show the resultant feature extraction signatures for Comms-only and Physical-only metric selections, and Fig 4, these metric spaces are brought together and reassessed.

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 2 and $INDD$ in the Selfish Target Selection behaviour in Fig 3). 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.

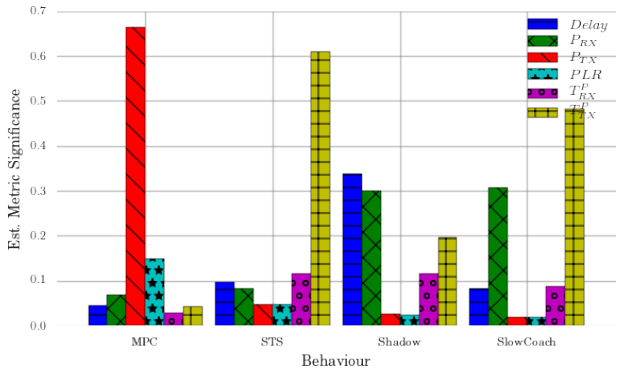

Figure 2: Plot of X_{comms} Metric Feature Extraction

5.2 Weight Assessment

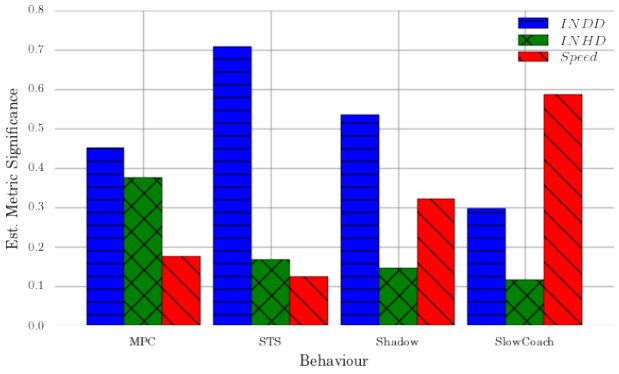

Figure 3: Plot of X_{phys} Metric Feature Extraction

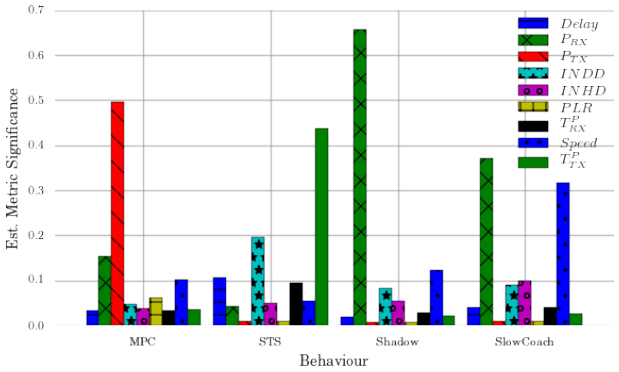

Figure 4: Multi Domain Relevance assessment of Metric Features

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 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 (6).

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 5-10,

Table 2: 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

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 5 and 6, 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 2 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 2 and 4, it's clear that the transmitted throughput (T^P_{TX}) 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 and 8, 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 2, the Shadow behavior is the most consistently detectible behaviour across domains.

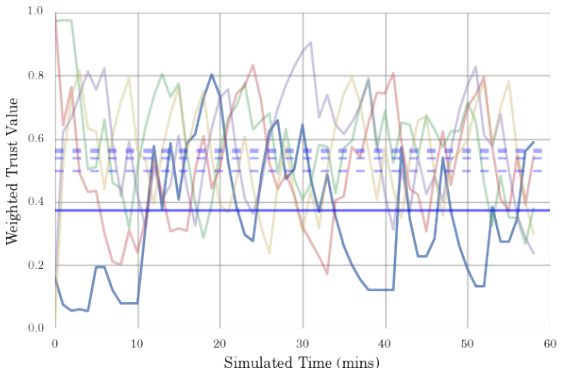

Figure 5: Selfish(STS) Targeting Comms Metric Trust

6. 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,

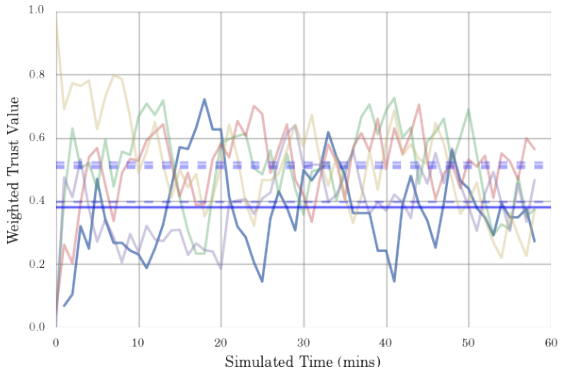

Figure 6: Selfish(STS) Targeting Full Metric Trust

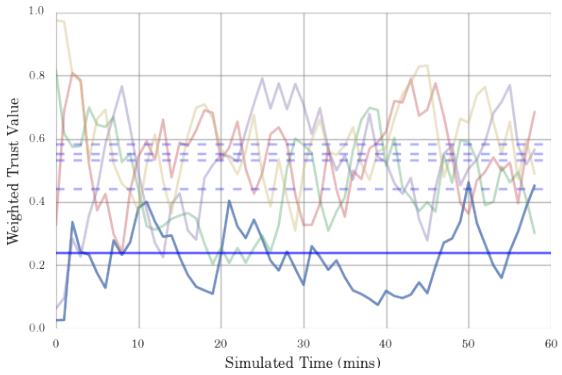

Figure 7: Shadow Comms Metric Trust

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.

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

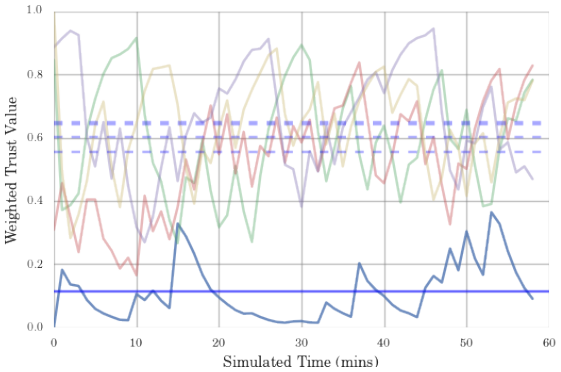

Figure 8: Shadow Full Metric Trust

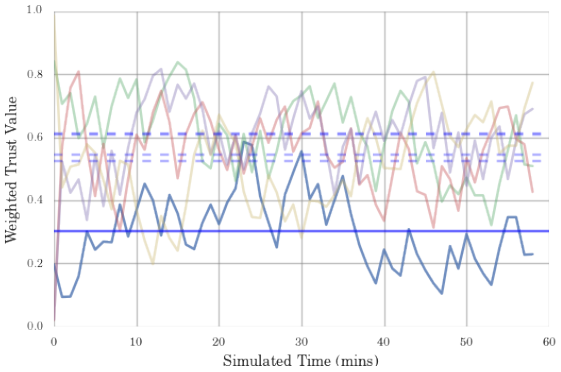

Figure 9: SlowCoach Comms Metric Trust

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.

REFERENCES

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

[2] L. Breiman. Random forests. *Mach. Learn.*, pages 5–32, 2001.

[3] S. Buchegger and J.-Y. 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.

[4] J.-h. Cho, A. Swami, and I.-r. Chen. A survey on trust management for mobile ad hoc networks. *Commun. Surv. & Tutorials*, 13(4):562–583, 2011.

[5] J. Guo. Trust and Misbehaviour Detection Strategies for Mobile Ad hoc Networks. 2012.

[6] J. Guo, A. Marshall, and B. 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.

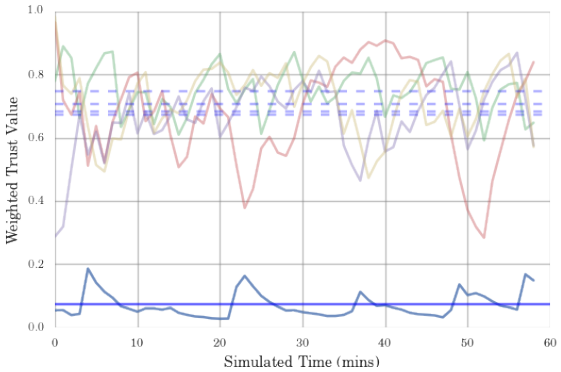

Figure 10: SlowCoach Full Metric Trust

[7] L. H. L. Hong, W. C. W. Chen, L. G. L. Gao, G. Z. G. Zhang, and C. F. C. 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.

[8] J. D. Lee and K. A. See. Trust in automation: designing for appropriate reliance. *Hum. Factors*, 46(1):50–80, 2004.

[9] H. Li and M. Singhal. Trust Management in Distributed Systems. *Computer (Long. Beach. Calif).*, 40(2):45–53, 2007.

[10] J. Li, R. Li, J. Kato, J. Li, P. Liu, and H.-H. Chen. Future Trust Management Framework for Mobile Ad Hoc Networks. *IEEE Commun. Mag.*, 46(4):108–114, apr 2007.

[11] J. Luo, X. Liu, Y. Zhang, D. Ye, and Z. Xu. Fuzzy trust recommendation based on collaborative filtering for mobile ad-hoc networks. *2008 33rd IEEE Conf. Local Comput. Networks*, pages 305–311, 2008.

[12] G. S. Malkin. RIP Version 2. STD 56, RFC Editor, nov 1998.

[13] J. Milgram, C. V. Alt, and T. Prestero. Verification of a Six-Degree of Freedom Simulation Model for the REMUS Autonomous Underwater Vehicle by in partial fulfillment of the requirements for the degrees of and at the Chairperson , Committee on Graduate Students Verification of a Six-Degree of F. 2001.

[14] J. Miquel and J. Montana. AUVNetSim: A Simulator for Underwater Acoustic Networks. *Program*, pages 1–13, 2008.

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

[16] K. Müller and T. Vignaux. SimPy: Simulating Systems in Python. *ONLamp.com Python DevCenter*, feb 2003.

[17] J. Partan, J. Kurose, and B. N. Levine. A survey of practical issues in underwater networks. *Proc. 1st ACM Int. Work. Underw. networks WUWNet 06*, 11(4):17, 2006.

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

[19] A. Stefanov and M. Stojanovic. Design and

performance analysis of underwater acoustic networks. *IEEE J. Sel. Areas Commun.*, 29(10):2012–2021, 2011.

[20] M. Stojanovic. On the relationship between capacity and distance in an underwater acoustic communication channel, 2007.

[21] C. Zouridaki, B. L. Mark, M. Hejmo, and R. 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.

[22] F. Zuo. Determining Method for Grey Relational Distinguished Coefficient. *SIGICE Bull.*, 20(3):22–28, jan 1995.