Physical Behaviours for Trust Assessment in Autonomous Underwater MANETs

Andrew Bolster, Alan Marshall
Department of Electrical Engineering and Electronics
University of Liverpool
Liverpool, UK
Email: andrew.bolster,alan.marshall@liverpool.ac.uk

***Abstract*—This paper proposes a new approach to determine trust in resource-constrained networks of autonomous systems based on their physical behaviour, using the motion of nodes within a team to detect and identify malicious or failing operation within their cohort. This is accomplished by looking at operations in the underwater marine environment. We present a series of composite metrics based on physical movement, and apply these metrics to the detection and discrimination of sample physical misbehaviours. This approach opens the possibility of bringing information about both the physical and communications behaviours of autonomous MANETs together to strengthen and expand the application of future Trust Management Frameworks in sparse and/or resource constrained environments.**

1. Introduction

Early attempts to secure and protect the integrity of Mobile Ad-hoc Networks have relied on various forms of strong-cryptography to protect information being transferred from tampering or malicious inspection. While such approaches protect the integrity of individual pieces of data, the increased computation, and storage requirements of modern, strong, decentralised cryptographic systems presents a clear avenue for Denial of Service (DoS) attacks on MANETs [1]. This threat is particularly relevant in resource-constrained networks, where one or more aspects of the environment are limited, be it available power, mobility, data storage, onboard processing, bandwidth, and channel resources such as capacity and delay. In such networks, where there is a requirement of security and/or integrity monitoring, strong-cryptographic methods present an entirely new opportunity to potential attackers.

One solution to the trade-off between DoS-protection, and security is the assessment of "trustworthiness" of nodes within a local network. "Trust" is an assessment of the capability of a node based on previously observed behaviour. Using this Trust to make simple routing decisions is significantly simpler and faster that strong-cryptographic methods, particularly in multi-hop networks or resource constrained networks [1]. With Trust being reliant on the near-real-time awareness of some behaviour, and cryptography on the pre-establishment of some entropy store and the repeated reinforcement of that numerical security, these represent two very different approaches to system integrity with very different costs/benefits. In practice, some elements of both methodologies will be used in different contexts and applications. These approaches to operational security have been totally focused on the establishment of trust/security in the communications domain, and ignore other potential threats to the network can be exploited through physical movement. This threat is particularly evident in collaborative autonomous systems where nodes are tasked to accomplish some survey / exploration / observation objective in a distributed fashion, where individual nodes make decisions based on the actions of their "team".

This collaboration opens the opportunity for a physically-misbehaving actor to selfishly conserve it's own resources, or maliciously "drain" a given target node. Current security / trust systems applied to MANETs are not concerned with the threat of such physical misbehaviours. This paper proposes a new approach to trust in resource-constrained networks of autonomous systems based on their physical behaviour. This paper proposes a new approach to trust in resource-constrained networks of autonomous systems based on their physical behaviour. This is accomplished by looking at operations in the underwater marine environment and using the motion parameters of nodes in a team to generate a series of composite metrics based on physical movement, and apply these metrics to the detection and discrimination of sample physical misbehaviours.

In the majority of Trusted autonomous mobile network implementations, a free space RF communications protocol such as 802.11 is used to derive all information about the trustworthy operation of the network. Most of these trust frameworks use a single type of observed communication action to derive trust assessments, typically successfully delivered or forwarded packets. By their nature, such implementations rely on relatively high bandwidth, low noise, low latency, and high channel occupancy where contention is tolerable. In contrast; in underwater environments, communications are sparse, delayful, noisy, and very prone to destructive contention. Observations of the communications processes used to assess trust 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, power and locomotive limitations and the increasing drive towards the use of teams of smaller, cheaper, almost disposable autonomous underwater vehicles (AUVs), particularly in defence, ecological and petrochemical fields, present unique threats against trust management.

In Section 2, we review the current use cases, deployments and mobility patterns of collaborative AUV operations, and the state-of-the-art in underwater localisation techniques. In Section 3, we discuss the use of TMFs and their applicability to marine operations. In Section 4, we propose a collection of metrics to characterise the physical behaviours of nodes, and establish a set of physical "misbehaviours" to assess these. In Section 5, we design a series of simulations, and tests to assess the detection and identification capabilities of three potential physical metrics for trust assessment. In Section 6, we assess the successful detection and identification characteristics of a series of tests, culminating in the generation and testing of a simple rule based behaviour classifier.

2. AUV Mobility and Localisation

The use of Autonomous Underwater Vehicles (AUVs) has greatly expanded in recent years; current applications and considerations are summarised below.

2.1. AUV operations and deployments

2.1.1. Hydrographic Survey. The use of AUVs in the place of manned-surface platforms or tethered undersea platforms enables greatly increased spatial and temporal sampling. Importantly, the separation of AUVs from the noisy sea surface enables much more efficient survey operations. This is particularly important when comparing to classical tow-line based measurements; where the mobility of the AUVs enables for much tighter-turning survey patterns or operation in inaccessible or hard-to-reach locations such as polar survey [2].

Another significant factor is cost; the daily cost of operating a manned vessel can be considerably higher than the costs of deploying, operating and recovering one or more AUVs with equivalent capabilities [3]. Additionally, the use of low-power "glider" AUVs has lowered the barrier to entry for extended mission types, such as persistent environmental survey, or open-ocean operations. Depth-hardened AUVs have also opened up the deepest parts of the oceans to exploration, with onboard autonomy, imagery and Simultaneous Location and Mapping (SLAM) techniques allowing deep-dwelling survey AUVs to react to bottom-surface features without the need for a tight craft-to-surface control loop [4]. The natural extension of these kind of applications is the use of AUVs on ice-covered planets and moons such as Europa, where three-dimensional, autonomous navigation without an on-the-loop controller is vital for mission resource efficiency and success.

2.1.2. Hull & Infrastructure Inspection. Concerns regarding the security, safety and legality of international shipping has driven the use of AUVs for near-surface hull and infrastructure inspections, looking for damage as well as devices such as limpet mines and contraband. This puts a range of unique pressures on the AUV system; requiring highly accurate three-dimensional localisation and path-planning to clearly image the contours of a hull [3]. With the increasing use and criticality of intercontinental undersea optical fibre connections, using AUVs for both the laying of and inspection of these cables is an exciting area of work [5][6].

2.1.3. Marine Petrochemical. Oil & Gas industry requirements for high quality, low altitude bathymetry of seabed structures for infrastructure development (pipelines/drill platforms etc.) as well as monitoring of those structures over time (inspection etc.) is another driver of research investment. As in Hydrography, the mobility of AUVs is the biggest single advantage over classical platforms[7].

2.1.4. Military. Mine-Countermeasure operations benefit greatly from, and significantly drive, AUV development; the ability to rapidly explore and covertly survey a potentially dangerous area without risking a human operator is a major benefit. This benefit applies to protection as well as incursion; the ability to have persistent survey of a valuable area such as a forward-operating harbour is increasingly essential, and as AUV technology, autonomy and security practices develop, this use is increasing. This Port Protection capability is particularly complex; teams of AUVs are expected to repeatedly survey an area and remain densely-connected enough to maintain end-to-end communications with all other nodes, in the face of an environment that is possibly not well surveyed initially, and includes dynamically moving obstacles (i.e. ships). In Sec. 5, we use this Port Protection scenario as a baseline for our simplified simulation context.

2.2. Localisation Technologies

Given the subsurface nature of most AUV operations, terrestrial localisation techniques such as GPS are unavailable (below $\approx 20cm$ depth). However, a range of alternative techniques are used to maintain spacial awareness to a high degree of accuracy in the underwater environment.

2.2.1. Long baseline (LBL). Long-baseline localisation systems use a series of (usually) static surface/cable networked acoustic transponders to provide coordinated beacons and (usually) GPS-backed relative location information to local subsurface users. Such systems can be accurate to less that $0.1m$ or better in ideal deployments and are regularly used in controlled autonomous survey environments such as harbour patrol operations where the deployment area is bounded. However, the initial setup and configuration required in advance of any AUV operation makes LBL difficult to utilise in unbounded or contended areas. LBL systems can also be deployed on mobile surface platforms in the area (ships or buoys for example), but these applications put significant computational pressure on the end-point AUV and have greatly reduced accuracy compared to ideal deployments [8].

2.2.2. Doppler Velocity Log (DVL). DVL uses the emission of directed acoustic "pings" that reflect off sea bed/surface interfaces which, when received back on the craft with multi-beam phased array acoustic transducers can measure both the absolute depth/altitude (z-axis) of the craft and through directional Doppler shifting, the relative (xy-translative) motion of the craft since the ping. While classical DVL was highly sensitive to shifting currents in the water column, advances in the development of Acoustic Doppler Current Profiling (ADCP) has turned that situation on its head, enabling the compensation-for and measurement-of water currents down to the sub-meter level [9].

2.2.3. Inertial Navigation Systems (INS). Inertial navigation systems use gyroscopic procession to observe the relative acceleration of a mobile platform. This reference-relative monitoring is particularly useful in the underwater environment, as it detects the motion of AUVs as they are carried by the water itself. Bias Drift is a significant problem for INS systems operating over longer (hundreds of metres) distances, as they usually have some minimal amount of directional bias which incurs a cumulative effect over time without correction. Several sensor synthesis processes have been demonstrated which combine information from INS along with DVL data to improve localisation into the sub-decimeter level [10].

2.2.4. Simultaneous Location and Mapping (SLAM). Simultaneous Location and Mapping is the process of iteratively developing a feature-based model of an environment, and to use the relative movement within that modeled environment to obtain estimates of absolute positioning. SLAM has been most well developed in the contexts of either visual-based inspection using cameras, or LIDAR-style distance triangulation, however the same principles have been successfully applied using marine sonar readings, providing sub-meter accuracy, real-time, feature-relative localisation information that is (for the most part) environmentally agnostic [11].

In summary, current technology enables AUVs to localise to a sub-metre accuracy in most contexts.

3. Trust Management Frameworks

Trust Management Frameworks (TMFs) provide information to assist the estimation of future states and actions of nodes operating as teams or networks. This information is used to optimize the performance of a team against malicious, selfish, or defective misbehaviour by one or more nodes. Previous research has established the advantages of implementing communications-based TMFs in terrestrial, 802.11 based MANETs, particularly in terms of preventing selfish operation in collaborative systems [12], and maintaining throughput in the presence of malicious actors [13]. These observations then inform future decisions of individual nodes, for example, route selection [14].

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) uses a range of communications metrics beyond packet loss rate (PLR) to assess trust [15]. This vectorized trust also allows a system to detect and identify the tactics being used to undermine trust. This method as been previously applied to the marine space, comparing against a selection of existing communications TMFs showing that MTFM is more effective at detecting misbehaviours in sparse communications environments [16].

4. Physical Behaviours for Trust

4.1. Physical Metrics

Three physical metrics are used 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. 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. As such, these metrics completely encapsulate and abstract the physical behaviour of any node, potentially performing any misbehaviour. Given that local nodes within the team are aware of the reported positions and velocities of their neighbours, it is believed that this is a reasonable initial set of metrics to establish the usefulness of physical metrics of trust assessment.

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

$$INHD_{i,j} = \hat{v}|v = V_j - \sum_x \frac{V_x}{N} \tag{2}$$

$$V_{i,j} = |V_j| \tag{3}$$

Where i and j are indices denoting the current observer node and the current observed node respectively; x is a summation index representing other nodes in the observers region of concern; Pj is the $[x, y, z]$ absolute position of the observed node (relative to some coordinated origin point agreed upon at launch) and Vj is the $[x, y, z]$ velocity of the observed node. Thus, the metric vector used for the physical-trust assessment from one observer node to a given target node is;

$$X_{i,j} = \{\text{INDD}_{i,j}, \text{INHD}_{i,j}, , V_{i,j}\} \tag{4}$$

At each time-step, each node will have a separate X assessment vector for each node it has observed in that time.

Additional metric sets may be more suitable for certain contexts, platforms or operations, however these were selected in collaboration with UK DSTL and NATO CMRE as suitable, generic, assessments, viable on most current platforms in most current deployment schemes.

4.2. Physical Misbehaviours

Misbehaviours in the communications space is heavily investigated area in MANETs [17][18][19], but attacks and misbehaviours in the physical space are far less explored. As in the communications space, the primary drivers of any deliberate "misbehaviour" come under two general categories; selfish operation or malicious subterfuge. Autonomous MANETs in general rely (or are at least, most effective) when all nodes operate fairly, be that in terms of their bandwidth sharing, energy usage, routing optimality or other factors. Physically, if a node is being "selfish", it may preferentially move to the edge of a network to minimise it's dynamic work allocation, or depending on it's intent, may insert itself into the centre of a network to maximise it's ability to capture, monitor, and manipulate traffic going across the network. In the context of a secure operation (or one that's assumed to be secure), the opportunity for capturing a legitimate node and replacing it with a modified clone. Assuming a highly capable outside actor and a multi-channel communications opportunity, there is also the possibility of a node appearing to "play along" with the crowd that occasionally breaks rank to route internal transmissions to a outside agent. In the underwater context this may mean an AUV following the rest of a team along a survey path and occasionally "breaking surface" to communicate to a malicious controller. Alternatively, if an inserted node is not totally aware of a given mission parameter, such as a particular survey or waypointing path, it may simply follow along, hoping not to be noticed.

In all these cases, such behaviour involves some element of behaving differently from the rest of the team, however, there are other cases where such individual "deviance" is observed; where a node is in some kind of "failure state". In the underwater context, this could be damage to the drive-train or navigation systems, causing it to lag behind or consistently drift off course. An ideal physical trust management system should be able to differentiate between "malicious" behaviours and "failing" behaviours.

To investigate this hypothesis, we create two "bad" behaviours; one deliberately malicious, where a cloned node is unaware of the missions' survey parameters and attempts to "hide" among the fleet, and a "failing" node, with an impaired drive train, increasing the drag force on the node's propulsion system (conceptually a simulated propeller-strike). These two behaviours are designated *Shadow* and *SlowCoach* respectively.

5. Simulation and Validation

5.1. Simulation Background

Simulations were conducted using a Python based agent framework, SimPy [20], with a network stack built upon AUVNetSim [21], with transmission parameters taken from and validated against [22] and [23]. For the purposes of this paper, this network is used for the dissemination of node location information, assuming suitable compression of internally assumed location data compressed into one 4096 bit acoustic data frame, with the network overall emitting approximately 10 frames a minute. Node kinematics are modelled on REMUS 100 AUVs, based on limits and core characteristics given in [24], [25] and [26]. These limits are given in Table 1. For the purposes of this exploratory case we do not model the hydrodynamics of the control surfaces of the AUVs, however we do model axial drag as a resistive inertial force.

TABLE 1. REMUS 100 MOBILITY CONSTRAINTS AS APPLIED IN SIMULATION

Parameter	Unit	Value
Length	m	5.5
Diameter	m	0.5
Mass	kg	37
Max Speed	ms^{-1}	2.5
Cruising Speed	ms^{-1}	1.5
Max X-axis Turn	$^{\circ}s^{-1}$	4.5
Max Y-axis Turn	$^{\circ}s^{-1}$	4.5
Max Z-axis Turn	$^{\circ}s^{-1}$	4.5
Axial Drag Coefficient (c_d)	NA	3
Cross Section Area	m^2	0.13

5.2. Node Control Modelling

We use the example of a Port Protection scenario, where a team of six AUVs are tasked with surveying a simplified harbour; a $1km$ x $1km$ x $100m$ cuboid volume. This is accomplished through a distributed waypoint system where by the team must regularly "check" several points around the exterior and interior of this volume. Boidean flocking behaviour [27] is used in addition to the cubic waypoint-survey behaviour to provide both collision-avoidance capability and maintaining node communications. This consists of three heuristic rules; Cohesion, Repulsion and Alignment.

$$F_{j,C} = F_{+}\left(p_j, \frac{1}{N}\sum_{\forall i\neq j}^{N} p_i, d_{max}\right) \tag{5}$$

$$F_{j,R} = \sum_{\forall i\neq j}^{N} F_{-}\left(p_j, p_i, d_{max}\right)\big| d_{max} > \|p_i - p_j\|\big) \tag{6}$$

$$F_{j,A} = \frac{1}{N}\cdot\left(\sum_{\forall i\neq j}^{N} \hat{v_i}\right) \tag{7}$$

Where F's are force-vectors applied to the internal guidance of the AUV, Cohesion $F_{j,C}$; Repulsion $F_{j,R}$; and Alignment $F_{j,A}$, F_{+} is a scaled vector attraction function, and F_{-} is an equivalent repulsion function

$$F_{+}(p^i, p^j) = (\widehat{p^i - p^j}) \times \frac{|p^i - p^j|}{d} = F_{-}(p^j, p^i) \tag{8}$$

5.3. Standards of Accuracy

The key question addressed in this paper is to assess the advantages and disadvantages of utilising trust from the physical domain. The "effectiveness" of any trust assessment framework is taken as consisting of several parts, the *accuracy* of detection and identification of a particular misbehaviour, the *complexity* of such analysis, including any specific training required, and the *differentiability* of behaviours using given metrics. In this case we are particularly interested in the accuracy of detection and identification of malicious / failing behaviours, and as such are looking at three key characteristics of accuracy; true detection accuracy; false positive rates; and misidentification rates.

As such we have three primary questions to answer to establish if these metrics are useful: How accurate are these metrics in being able to easily differentiate between Normal and Abnormal behaviours in terms of True-Positive and False-Positive rates? What differentiation of metric response, if any, is there between the stated abnormal behaviours? Can a simple classifier be built to characterise these differentiations of response, and what is it's True-Positive/False-Positive accuracy?

5.4. Analysis

Sixty-four simulation runs are executed for each scenario (i.e. one node "Maliciously" following the fleet with no mission information (Shadow), one "Failing" node with simulated drive train issues (SlowCoach), and one baseline control scenario where all nodes are behaving appropriately (Control). Each of these simulated missions last for an hour, matching realistic deployment times based on current MOD/NATO operations[28].

In order to assess the viability of using the previously discussed metrics, the raw motion paths recorded by the simulation are fed into an analysis pipeline aimed at abstracting the instantaneous observed values into derived deviations from "normal" behaviour in the team.

$$d_{i,j}^{m,t} = x_{i,j}^{m,t} - \frac{\sum_k x_{i,k}^{m,t}}{|M|} \tag{9}$$

$$\alpha_{i,j}^{m,t} = \left|\frac{d_{i,j}^{m,t}}{\sigma(d_{i,j}^{m,t})}\right| \tag{10}$$

$$C_i^m = \sum_t \alpha_i^{m,t} * \left(\frac{\sum_{x\neq i}\Sigma_t \alpha_x^{m,t}}{N-1}\right)^{-1} \tag{11}$$

Where i and j are indices denoting the current observer node and the current observed node respectively; xk is a summation index representing other nodes in the observers region of concern; X is the vector of metrics from (4) and m is the index of a particular metric in X; d is an intermediate value of the deviance of a given observation from the mean, and α is a normalised response value in terms of it's deviation from the mean at that instance. C_i^m is an inferred "Confidence" value denoting the relative "Deviation of Deviation" between a given nodes response in a given metric to the rest of its cohort in that metric.

5.4.1. Behaviour Detection and Classification. A simple misbehaviour detection is to apply Dixon's Q-test [29] to the resultant $\sum_t \alpha$ values for each node for each metric for each run, establishing if a misbehaving node exists in a given run, and to identify that misbehaving node. We use a Confidence Interval of 95%. Our initial hypothesis is that by using observations of the previously stated physical metrics, that we will be able to detect and identify misbehaviours. Within that context, this Confidence Interval indicates that we would expect only a 5% chance that any run or node identified using the Q-test to *not* be a misbehaving run/node. Additionally, by applying the Q-test on a per-metric basis, we can use the "votes" of each metric as a simplified consensus classifier. This classifier may allow us to characterise some aspect of a given misbehaviour in terms of metrics it heavily impacts, and those that are less affected, finding some differentiating-limit between certain behaviours using certain metrics.

5.4.2. Operational Performance Metrics. While not the focus of this paper, we are also concerned with the impact of these misbehaviours on the mission efficiency. We monitor this in three main measurements; the speed of the fleet in terms of how many of it's port-protection waypoints nodes pass, the total energy used for communications, and the average end-to-end delay in the acoustic network. We would expect that any misbehaviour in positioning will incur some loss of efficiency, whether it is the fleet being slowed down by a "SlowCoach" attempting to catch up or of a node moving in an unexpected way dragging the team temporarily off course. Given that in acoustic communications, transmission is energetically expensive while reception is not, and while physical misbehaviours will not impact the amount of offered load on the network, collisions induced by un-even distribution of nodes should have a small but measurable effect on the energy used for packet reception.

6. Results and Discussion

Fig. 1 shows the raw metric values (vertically) from one run of each behaviour (horizontally). It clear from the INDD and INHD metric responses in the both misbehaviour cases (Shadow / SlowCoach), Alfa is the outlier and other nodes are all consistent in their metric values. This outlier-response is not nearly as clear in the Speed metric case (bottom row of Fig. 1). Looking at the differen behaviours; it appears that the Shadow behaviour is creating the largest, most obvious deviations. In Fig. 2 the metric values are normalised as per (10). This highlights the outlying-characteristic of INHD and INDD; largely eliminating the other nodes-responses. In the Speed response of Fig. 2, the Speed metric is not obviously highlighting any significant misbehaviours in that metric. From Fig. 3, normalising across the duration of each run, it appears that Speed is being affected differently between the two misbehaviours, and much less so than INHD/INDD.

6.1. Detection of Misbehaviours

We have demonstrated that INHD and INDD appear to accurately and obviously identify the malicious node in the case that there is one. Using the deviance normalisation presented in (10), we can observe clear, almost contiguous areas under the Alfa-values in Fig. 2 in the Shadow and SlowCoach misbehaviours. From Fig. 3, while the deviance in Speed is not as strong as the deviance in INHD and INDD, it varys between misbehaviours, indiciating that Speed may be a way to analytically differentiate between the two behaviours.

To investigate how this would relate to the ability to blindly detect misbehaviours, the Q-test is applied to $\Sigma\alpha$ results as used in Fig. 3, to attempt to correctly establish if a node is misbehaving and and if so, which node. As such the "correctness" rule for assessing this strategy is that, in misbehaving cases, the test should return "Alfa" (otherwise a "Fail" is recorded), and in the Control case, the test should assert that there are no outliers, (otherwise a "Fail" is recorded again). In Table 2, the Control case is correctly identified 92% of the time. The "malicious", Shadow misbehaviour is detected and identified 98% of the time, and the "failing", SlowCoach misbehaviour is identified just 79% of the time. These values match our intuition from Figs. 1 & 2.

TABLE 2. Overall Q-Test Outlier Detection Accuracy

Behaviour	Mean	Std
Control	0.927	0.261
Shadow	0.979	0.144
SlowCoach	0.792	0.408

TABLE 3. Per-Metric Q-Test Outlier Detection Accuracy

	Behaviour	INDD	INHD	Speed
Mean	Control	0.875	0.938	0.969
	Shadow	1.000	1.000	0.938
	SlowCoach	1.000	1.000	0.375
Std	Control	0.336	0.246	0.177
	Shadow	0.000	0.000	0.246
	SlowCoach	0.000	0.000	0.492

We can investigate this further by looking at the "correctness" of the assessments of each metric individually (Table 3). From this we can see that in both misbehaviours, INHD and INDD correctly identify Alfa as the misbehaver 100% of the time. However, they mis-detect a potential misbehaviour in the Control case 13% and 7% of the time respectively. Meanwhile, Speed correctly identified the Control case 97% of the time, and the Shadow case 94% of the time, but missed the SlowCoach behaviour 63% of the time. This result is surprising on the face of it, as SlowCoach is a misbehaviour that is exclusively about individual node speed and conceptually should have had a much larger impact on the simple Speed metric. However, the collaborative nature of the collision avoidance system, and the existing limits on node kinematics from Table 1 appear to be hiding this impact.

6.2. Identification of Misbehaviours

Having established the detection of physical misbehaviour to a statistically significant level, and a demonstrable difference in metric-response to different misbehaviours, we now focus on our last question from Sec. 5.3; can we construct a simple classifier based on a subset of our results and apply it blindly to a new set of results?

TABLE 4. Metric Confidence Responses

	Behaviour	INDD	INHD	Speed
Mean	Control	1.064	0.966	1.010
	Shadow	4.059	3.374	2.098
	SlowCoach	4.246	3.352	1.491
Std	Control	0.262	0.113	0.132
	Shadow	0.398	0.436	0.206
	SlowCoach	0.198	0.288	0.180

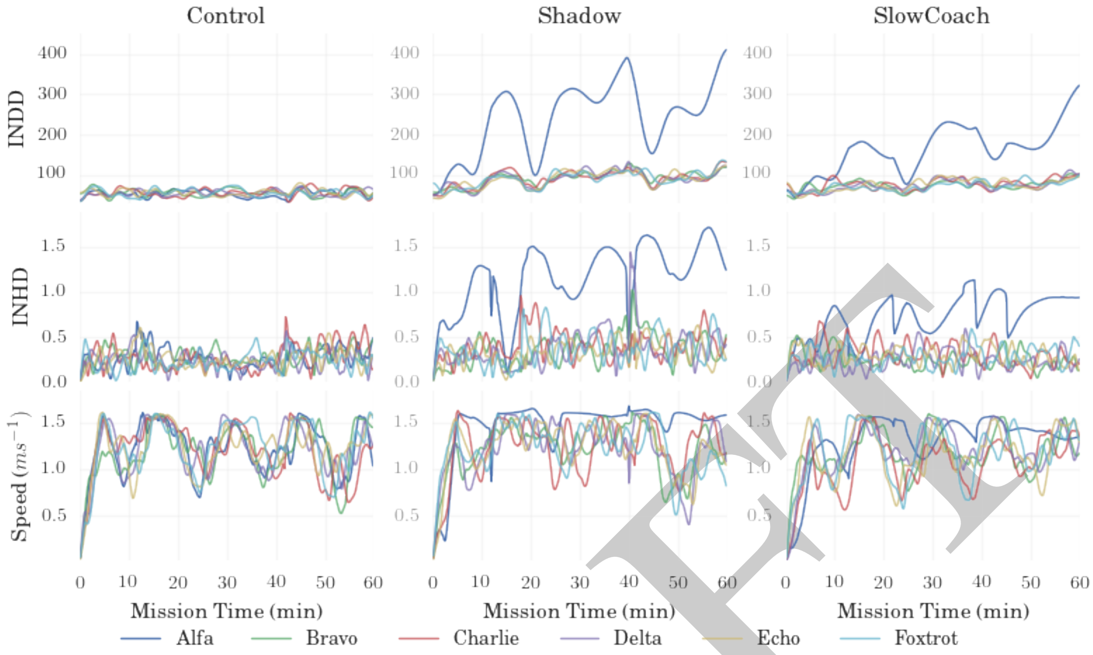

Figure 1. Observed Metric Values for one simulation of each behaviour ($x_{i,j}^{m,t}$ from (9))

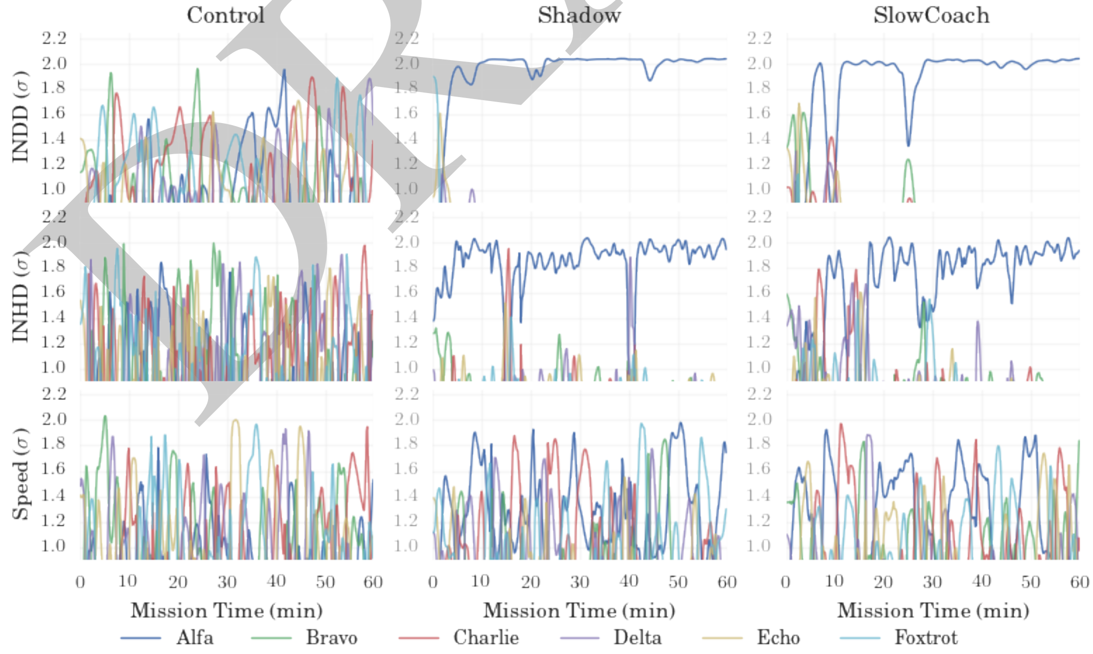

Figure 2. Normalised Deviance values from one simulation of each behaviour ($\alpha_{i,j}^{m,t}$ from (10))

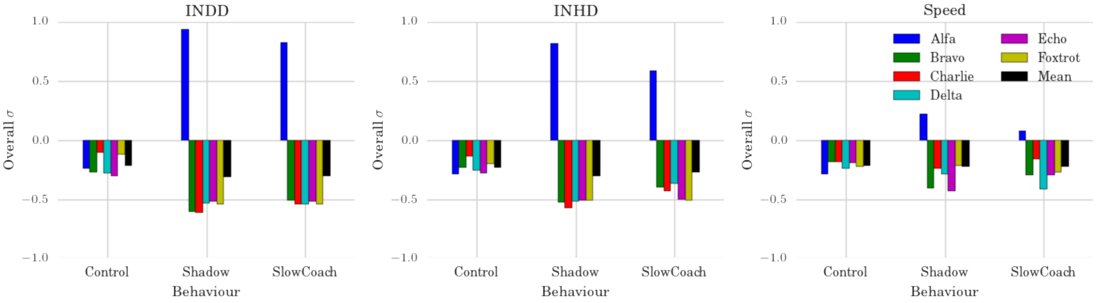

Figure 3. Per-Node-Per-Run deviance for each metric, normalised in time ($\sum \alpha/T$)

As we can already establish the Control case accurately, we continue to use the Q-test across all metrics for that case and concentrate of differentiating between the misbehaviours.

From (11), we can establish the per-metric-per-behaviour "Confidence" in the relationship between a given metric deviance and each behaviour and use the as a "signature" for that behaviour. From Table 4 we observe that INHD and INDD have similar responses to both misbehaviours, with significant standard deviations, but the response of the Speed metric is much more stable and discernible; across the range of training simulation runs, the SlowCoach behaviour centres around 1.5, while the Shadow behaviour centres around 2.0, with these centres being at least one standard deviation away from each other. Our generated classifier is formalised in (12).

$$C \rightarrow \begin{cases} Q^{95}(X) = \emptyset, & \text{Control} \\ Q^{95}(X) \neq \emptyset \wedge \text{Speed}^{X} \leq 1.75, & \text{Shadow} \\ Q^{95}(X) \neq \emptyset \wedge \text{Speed}^{X} > 1.75, & \text{SlowCoach} \end{cases} \tag{12}$$

Applying this simplified classifier to a blind test set of simulations (of the same scale) gives surprisingly positive results as shown in Table 5, with greater than 90% identification rates for both misbehaviours. However, in the Null (Control) case we experience a false-positive rate of nearly 30%, that is to say that in the case where there is no misbehaviour, 30% of the time a node will be mis-identified as misbehaving when it is not. These are strongly positive results for the use of physical metrics for behaviour discrimination; with INHD and INDD proving as strong and obvious "canaries" of misbehaviour, and Speed in this case proving a capable differentiator between conceptually close misbehaviours.

6.3. Impacts of Misbehaviour on operational performance

Anticipated "small but measurable" effects to communications performance and energy usage are extremely small and within the bounds of statistical uncertainty. One observation of merit was an observed 10% increase in end-to-end delay in the case of the Shadow behaviour, due to the misbehaving node "overshooting" the mission waypoints, losing connection to some more distant nodes, causing retransmissions and delays. Waypoint passing rates were identical to within 2% error across all behaviours, and fleet distance remained within a similar margin. It's possible that our selected behaviours were too unambitious in our impacts, and future work will have to investigate the impact of "heavy-handed" or destructive behaviours on the operational efficiency of autonomous networks.

TABLE 5. IDENTIFICATION RATES ON UNTRAINED RESULTS USING (12)

True Behaviour	Probability of Correct Blind Identification
Control	0.719
Shadow	0.906
SlowCoach	0.938

7. Conclusion

In this paper we have demonstrated that with current underwater localisation techniques, that in certain mobility models, that a set of geometric abstractions (INHD, INDD, and Speed), between nodes as part of an Underwater MANET can be used as a Trust Assessment and Establishment metric. These metrics are application-agnostic and could potentially be applied in other areas of mobile autonomy such as terrestrial, aerial, and mixed MANETs.

We show, using a Port-Protection waypoint-led scenario built upon a Boidian collision prevention behaviour that in a simulated underwater environment, the outputs of these metrics can be used to detect and differentiate between exemplar malicious behaviour and potential failure states. This verification further supports the assertions the assertion that it is practical to extend Trust protocols such as Multi-parameter Trust Framework for MANETS (MTFM) [15] to include metrics and observations from the physical domain as well as those from the communication domain [30]. This combination of physical and “logical” information further supports the decentralised and distributed establishment of observation based Trust.

Acknowledgment

The Authors would like to thank the DSTL/DGA UK/FR PhD Programme for their support during this project, as well as NATO CMRE for their advice and assistance.

References

[1] J. Cordasco and S. Wetzel, “Cryptographic Versus Trust-based Methods for MANET Routing Security,” *Electron. Notes Theor. Comput. Sci.*, vol. 197, no. 2, pp. 131–140, 2008.

[2] T. B. Curtin, J. G. Bellingham, J. Catipovic, and D. Webb, “AUTONOMOUS OCEANOGRAPHIC SAMPLING NETWORKS,” *Oceanography*, vol. 6, no. 3, pp. 86–94, 1993.

[3] J. Nicholson and A. Healey, “Underwater Acoustic Communications and Networking: Recent Advances and Future Challenges,” *Mar. Technol. Soc. J.*, vol. 42, no. 1, pp. 103–116, 2008. [Online]. Available: http://qub.library.ingentaconnect.com/content/mts/mtsj/2008/00000042/00000001/art00008

[4] L. Chen and H. Hu, “Towards Localization and Mapping of Autonomous Underwater Vehicles: A Survey,” 2011. [Online]. Available: http://cswww.sx.ac.uk/staff/hhu/Papers/CES-515AUVssurvey.pdf

[5] S.-C. Yu and T. Ura, “A System of Multi-AUV Interlinked With a Smart Cable For Autonomous Inspection of Underwater Structures,” *Int. J. Offshore Polar Eng.*, vol. 14, no. 04, 2004.

[6] K. Asakawa, J. Kojima, Y. Kato, S. Matsumoto, N. Kato, T. Asai, and T. Iso, “Design concept and experimental results of the autonomous underwater vehicle {AQUA EXPLORER} 2 for the inspection of underwater cables,” *Adv. Robot.*, vol. 16, no. 1, pp. 27–42, jan 2002. [Online]. Available: http://dx.doi.org/10.1163/156855302317413727

[7] B. Morr, “All Quiet on the AUV Front,” *Underw. Mag.*, no. February, pp. 1–5, 2003.

[8] a. Matos, N. Cruz, a. Martins, and F. L. Pereira, “Development and implementation of a low-cost LBL navigation system\nfor an AUV,” *Ocean. ’99. MTS/IEEE. Rid. Crest into 21st Century. Conf. Exhib. Conf. Proc. (IEEE Cat. No.99CH37008)*, vol. 2, pp. 774–779, 1999.

[9] J. Snyder, “Doppler Velocity Log (DVL) navigation for observation-class ROVs,” *MTS/IEEE Seattle, Ocean. 2010*, no. Dvl, pp. 1–9, 2010.

[10] X. Liu, X. Xu, Y. Liu, and L. Wang, “Kalman filter for cross-noise in the integration of SINS and DVL,” *Math. Probl. Eng.*, vol. 2014, no. Dvl, 2014.

[11] S. B. Williams, P. Newman, G. Dissanayake, and H. Durrant-Whyte, “Autonomous underwater simultaneous localisation and map building,” *Robot. Autom. 2000. Proceedings. ICRA ’00. IEEE Int. Conf.*, vol. 2, pp. 1793–1798, 2000.

[12] H. Li and M. Singhal, “Trust Management in Distributed Systems,” *Computer (Long. Beach. Calif).*, vol. 40, no. 2, pp. 45–53, 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4085622

[13] 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*. ACM Press, 2002, pp. 226–236. [Online]. Available: http://dl.acm.org/citation.cfm?id=513800.513828

[14] 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.*, vol. 46, no. 4, pp. 108–114, apr 2007. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=4212452http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4481349

[15] J. Guo, “Trust and Misbehaviour Detection Strategies for Mobile Ad hoc Networks,” 2012.

[16] A. Bolster and A. Marshall, “Single and Multi-metric Trust Management Frameworks for Use in Underwater Autonomous Networks,” in *Trust. 2015 IEEE*, vol. 1, aug 2015, pp. 685–693. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=7345343

[17] K. Konate and A. Gaye, “Attacks Analysis in Mobile Ad Hoc Networks: Modeling and Simulation,” *2011 Second Int. Conf. Intell. Syst. Model. Simul.*, pp. 367–372, jan 2011. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true{\&}arnumber=5730376{\&}contentType=Conference+Publications

[18] X. Wang, J. S. Wong, F. Stanley, and S. Basu, “Cross-Layer Based Anomaly Detection in Wireless Mesh Networks,” *2009 Ninth Annu. Int. Symp. Appl. Internet*, pp. 9–15, jul 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5230665

[19] R. Mitchell, I.-r. Chen, and V. Tech, “A Survey of Intrusion Detection in Wireless Network Applications,” 2014.

[20] K. Müller and T. Vignaux, “SimPy: Simulating Systems in Python,” *ONLamp.com Python DevCenter*, feb 2003. [Online]. Available: http://www.onlamp.com/pub/a/python/2003/02/27/simpy.html?page=2

[21] J. Miquel and J. Montana, “AUVNetSim: A Simulator for Underwater Acoustic Networks,” *Program*, pp. 1–13, 2008. [Online]. Available: http://users.ece.gatech.edu/jmjm3/publications/auvnetsim.pdf

[22] M. Stojanovic, “On the relationship between capacity and distance in an underwater acoustic communication channel,” p. 34, 2007. [Online]. Available: http://www.mit.edu/{∼}millitsa/resources/pdfs/bwdx.pdf

[23] A. Stefanov and M. Stojanovic, “Design and performance analysis of underwater acoustic networks,” *IEEE J. Sel. Areas Commun.*, vol. 29, no. 10, pp. 2012–2021, 2011.

[24] R. McEwen and K. Streitlien, “Modeling and control of a variable-length auv,” *Proc 12th UUST*, pp. 1–42, 2006. [Online]. Available: http://www.mbari.org/staff/rob/uustrep.pdf

[25] 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.

[26] S. A. Samad, S. K. Shenoy, G. S. Kumar, and P. R. S. Pillai, “A Survey of Modeling and Simulation Tools for Underwater Acoustic Sensor Networks,” *Networks*, pp. 40–47, 2011.

[27] C. W. Reynolds, “Boids (Flocks, Herds, and Schools: a Distributed Behavioral Model),” *SIGGRAPH 87 Proc. 14th Annu. Conf. Comput. Graph. Interact. Tech.*, vol. 21, no. 4, pp. 25–34, aug 1987. [Online]. Available: http://dl.acm.org/citation.cfm?id=37402.37406http://www.red3d.com/cwr/boids/

[28] A. Bolster, “Analysis of Trust Interfaces in Autonomous and Semi-Autonomous Collaborative MHPC Operations,” The Technical Cooperation Program, Tech. Rep., 2014.

[29] R. B. Dean and W. J. Dixon, “Simplified Statistics for Small Numbers of Observations,” *Anal. Chem.*, vol. 23, no. 4, pp. 636–638, 1951. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/ac60052a025

[30] A. Bolster and A. Marshall, “A Multi-Vector Trust Framework for Autonomous Systems,” in *2014 AAAI Spring Symp. Ser.*, Stanford, CA, 2014, pp. 17–19. [Online]. Available: http://www.aaai.org/ocs/index.php/SSS/SSS14/paper/viewFile/7697/7724