Analytical Metric Weight Generation for Multi-Domain Trust in Autonomous Underwater MANETs

Invited Paper for Underwater Communications Security session

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

***Abstract*—Trust Management Frameworks (TMFs) are being used to improve the efficiency, security, and reliability of decentralized and distributed autonomous MANETs using metrics garnered from the communications activities of nodes within the networks. However, these do not perform well in sparse / harsh environments such as those found in Underwater Acoustic Networks (UANs) [1]. As node capabilities increase, the physical motion of nodes represent an additional domain of knowledge about the operations and behaviours of the network.**

In this paper we present a Machine Learning supported methodology for optimising metric weight vector generation, using metrics from both physical and communications domains to detect and identify a range of misbehaviours, demonstrating that by utilising information from multiple domains, trust assessment can be more sensitive and accurate than in single-domain (communications) assessment.

I. Introduction

Trust Management Frameworks (TMFs) in terrestrial MANETs has been an area of active research for many years; with implicitly limited resources in terms of power, mobility, and communications range, the elimination or isolation of malicious nodes within the MANET is an obvious method for maintaining performance and security. TMFs provide information to assist the estimation of future states and actions of nodes within networks. This information is used to optimise the performance of a network against malicious, selfish or defective misbehaviour by one or more nodes. Existing research has demonstrated the advantages of implementing TMFs to 802.11 based MANETs in terms of preventing selfish operation and maintinaing throughput in the presence of malicious nodes [2], [3]. However, as these decentralised networks expand beyond the terrestrial arena into aerial and underwater environments, these Trust frameworks must be assessed for their suitability to these new regions of operation.

With respect to the comparatively stable terrestrial RF environment, the underwater acoustic communications environment is unforgiving; generally static or stable assumptions about propagation delay and paths, frequency-based attenuation and minimal refraction simply don't apply to many marine communications channels. This, coupled with the highly location, depth, weather, and flora and fauna dependent variability of these fundamental channel parameters make it difficult to make strong assumptions about if instantaneous network performance is the fault of a misbehaving node or of a whale passing between nodes.

Having previously established the usefulness of Physical Metrics (those based on the node movements and behaviours) in trust assessment, these metrics present the opportunity to use additional information to establish and maintain trust assessment [4].

In this paper we build upon previous work [1] that demonstrated the use of Random Forest regression [5] to assess the relative importance of Communications Metrics in a simulated UAN, and extend the presented methodology with novel optimisation target functions and applying these methods of Metric Assessment across both Physical and Communications metric domains.

This paper is laid out as follows: Sec. II outlines the operation and parameters of MTFM and summarises the comparison of MTFM and other classically terrestrial MANET TMFs in a simulated marine environment. Sec. III discusses the proposed weight generation and assessment scheme for Multi-Domain MTFM, presents the experimental scenarios used, and the analysis method applied. The results of this analysis are discussed in Sec. IV.

II. Trust for Marine Communication

A. Terrestrial Trust Management Frameworks

1) Single Metric TMFs: Single Metric TMFs such as Hermes [6], Objective Trust Management Framework (OTMF) [7] and CONFIDANT [3] can be generalised as single-value estimation based on a binary input state (success or failure of packet delivery); generating a probabilistic estimation of the future states of that input.

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. 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.

2) Multi-parameter Trust for MANETs (MTFM): Guo et al. [8] demonstrated the ability of grey relational analysis (GRA) [9] 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. MTFM performs cohort based normalization of metrics at runtime, providing a grey "grade" of trust compared to other observed nodes in that interval, while maintaining the ability to abstract trust values for decision support without requiring per-environment calibration or characterisation. These assessments are relative in both fairly and unfairly operating networks; all nodes 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.

The grey relational vector is given as

$$\theta^t_{k,j} = \frac{\min_k |a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|}{|a^t_{k,j} - g^t_j| + \rho \max_k |a^t_{k,j} - g^t_j|} \quad (1)$$
$$\phi^t_{k,j} = \frac{\min_k |a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|}{|a^t_{k,j} - b^t_j| + \rho \max_k |a^t_{k,j} - b^t_j|}$$

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. The implications of this are discussed in Sec. III).

This Grey Vector is weighted on a per-metric basis(2) to then generate a scalar trust value (3).

$$[\theta^t_k, \phi^t_k] = \left[\sum_{j=0}^{M} h_j \theta^t_{k,j}, \sum_{j=0}^{M} h_j \phi^t_{k,j}\right] \quad (2)$$
$$T^t_k = (1 + (\phi^t_k)^2/(\theta^t_k)^2)^{-1} \quad (3)$$

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 (1) the $[\theta, \phi]$ values are reduced into a scalar trust value by(3)[10]. MTFM combines this GRC with a topology-aware weighting scheme and a fuzzy whitenization model, however the model parameter we are primarily interested in is the weighting of metrics for MTFM, which enables the detection and identification of misbehaviours.

B. Summary of Previous Work

In [1], MTFM was compared against Hermes and OTMF in a series of simulated UANs. This was framed against a similar comparison applied to validate the development of MTFM [11] in terrestrial applications, and as such required rate and range scaling for application to the simulated underwater channel.

This comparison demonstrated that while the performance of MTFM, Hermes and OTMF were all severely reduced in the marine environment, Unweighted MTFM showed a consistent if small (10%) deviation in misbehaviour-cases and no such deviation in the Fair case. When the individual metrics in MTFM were preferentially weighted, it was demonstrated that a "Relevance Signature" could be generated for detectable behaviours based on a Random Forest regression of multiple weighted assessments, producing weighting vectors to maximally detect and identify the operation of those misbehaviours.

In [4], it was demonstrated that physical metrics and related physical misbehaviours could also be used to establish a simplified misbehaviour detection and identification classifier with a high degree of accuracy and selectivity ($> 90\%$ accurate identification of misbehaviours with $> 30\%$ false positive rate for control behaviours) based on a simple statistical outlier detection test (Dixons Q-test [12])

III. Weight assessment schemes for MTFM

From (2), 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. The construction of the g and b vectors from 1 depends on the trustworthiness-correlations of a particular metric, e.g. Throughput is assumed to be positively correlated to trustworthiness and so follows the basic construction ($g \mapsto \max, b \mapsto \min$). However, in the case of a metric such as delay, this relationship is inverted; longer delays indicate less trustworthy activity. In complex environments, the relationship between a metrics trustworthiness correlation may not be quite so obvious as the throughput / delay examples. This was recognised by Guo, but this "emphasis/correlation" weighting was manually configured for each metric for each behaviour and no quantative method for establishing such relationships had been presented since.

We proposed a purely computational methodology for weight generation using a Random Forest Regression and successfully applied it to the communications domain [1]. We now apply and extend this method to the joint physical-communications metric space.

It is important to establish what analytical behaviour is being targeted; in this case this target is the identifiable exposure of a "low" trust value for a misbehaving node that is as distinct from other "fair" nodes as possible across a full (or multiple) simulation runs of a particular behaviour given a particular weight vector. We characterise this "objective function" as ΔT_{ix} (4)

$$\Delta T_{ix} = \frac{\sum_{j \neq x} \left(\overline{T_{i,j}}^{\forall t}\right)}{N - 1} - \overline{T_{i,x}}^{\forall t} \quad (4)$$

(5)

Where i is a given observer, x is the suspected misbehaving node, $\overline{T_{i,j}}^{\forall t}$ is the average weighted trust assessment from 3

of node j observed by node i across time and N is the number of nodes in the current cohort.

Conceptually, ΔT_{ix} is the "Distrust" of the target node x, as the difference in trust value from $0 \rightarrow 1$, the higher the better for our purpose. As such, we aim to maximise $\Delta T_{ix} \forall i \neq x$.

A. Available Domain Metrics

1) Communications Metrics: We use the same trust metrics from [11] that are applicable to the marine environment, i.e. Delay, Received and Transmitted power, Throughput (S), Offered Load (G), and Packet Loss Rate (PLR). Thus, the metric vector for communications-trust assessment is;

$$X_{comms} = \{D, P_{RX}, P_{TX}, S, G, PLR\} \tag{6}$$

2) Physical 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 [4]

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.

$$INDD_{i,j} = \frac{|P_j - \sum_x \frac{P_x}{N}|}{\frac{1}{N}\sum_x \sum_y |P_x - P_y| (\forall x \neq y)} \tag{7}$$

$$INHD_{i,j} = \hat{v} | v = V_j - \sum_x \frac{V_x}{N} \tag{8}$$

$$V_{i,j} = |V_j| \tag{9}$$

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

$$X_{phy} = \{\text{INDD}, \text{INHD}, V\} \tag{10}$$

3) Cross Domain Trust Metrics: This simplest possible combination is a vector concatenation across domain metric vectors;

$$X_{merge} = (X_{comms} | X_{phy}) \tag{11}$$

$$= \{D, P_{RX}, P_{TX}, S, G, PLR, \text{INDD}, \text{INHD}, V\} \tag{12}$$

B. Simulation and Scenario Generation

To investigate the operation of a fully mobile network of six nodes, each kinematically modelled on the commonly used REMUS 100 AUV platform [13] in the marine environment, simulations were conducted using a Python based framework, SimPy [14], with a network stack built upon AUVNetSim [15], using transmission parameters taken from and validated against [16] and [17].

Four scenarios were developed to assess both communications and physical domains where one node within the fleet was misbehaving (n_m, normally designated "Alfa"). 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 cooperating less with the rest of the team through, while n_t 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 the fleet.
4) Slow Coach, where n_m is experiencing a simulated drive-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 port-protection / survey mobility pattern with Boidean collision avoidance control [18]. We perform 16 randomly seeded repetitions for each of the four misbehaviour cases, MPC, STS, Shadow, Slow Coach, and 32 repetitions of the default "Fair" behaviour.

C. Metric Weight Analysis Scheme

With the nine selected metrics of X_{merge}, 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, resulting in 18661 unique normalised weight combinations. These weights are applied to the simulated results, producing a $N \times N$ ΔT_{ix} matrix for each run for each behaviour for each weight.

We apply a Random Forest regression [5] to assess the relative importance of the selected metrics on increasing ΔT_{ix}. Random Forest accomplishes this by generating a large number of random regression trees and prunes these to fit incoming data. A major advantage of Random Forest is that we acquire an already normalised "relevance" vector mapping to the input weights for the particular behaviour comparison being tested.

After establishing the importance of weights in identifying particular behaviours, a final weight is arrived at by pruning those metrics that are unimportant (i.e. taking the three most relevant metrics in each behaviour) and iteratively swapping the g and b vector elements for those metrics to establish if the "trustworthiness" correlations are positive or negative. Finally, these resultant weights (e.g. Table I) are applied to a new, untrained, simulation set of the same size to generate new ΔT_{ix} values. Using this approach we can explore the results of many simulations, condensing the multi-dimensional problem (target / observer / behaviour / metric / time) to a more tangible level for analysis. We apply this method to both domains independently (X_{comms} and X_{phys}) to compare the performance of each metric set independently.

TABLE I
MULTI DOMAIN (X_{merge}) OPTIMISED WEIGHT VECTORS

Behaviour	Behaviour			
	MPC	STS	Shadow	SlowCoach
$Delay$	-0.187	-0.195	0.004	-0.157
P_{RX}	0.129	-0.035	-0.654	-0.533
P_{TX}	0.579	0.019	0.030	0.013
S	0.006	-0.100	-0.016	-0.132
PLR	0.069	0.019	0.030	0.013
G	-0.146	0.381	0.063	-0.028
$INDD$	0.040	-0.209	0.120	0.159
$INHD$	-0.190	0.057	0.158	0.206
$Speed$	-0.297	0.062	0.266	0.460

TABLE II
ΔT ACROSS DOMAINS AND DETECTED BEHAVIOURS

Domain	Behaviour				Avg.
	MPC	STS	Shadow	SlowCoach	
X_{merge}	0.90	0.10	0.50	0.63	0.53
X_{comms}	0.95	0.17	0.28	0.27	0.42
X_{phys}	0.02	0.02	0.43	0.76	0.31
Avg.	0.67	0.10	0.41	0.56	0.44

IV. RESULTS AND DISCUSSION

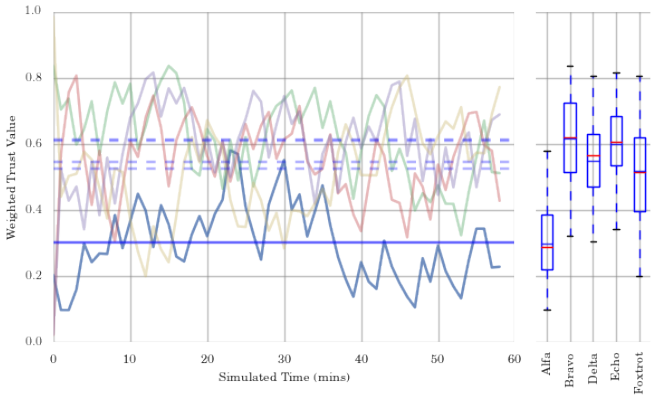
Looking first at the arrived-at weighted metrics for the Multi-Domain (X_{merge}) case; shown in Table I; we see that the "Physical" Shadow misbehaviour is most heavily (and inversely) weighted to received signal strength; as is the SlowCoach physical behaviour. To demonstrate the operation of this method, it is useful to look at the difference between the trust response using a single-domain-optimised weighting and that of a multi-domain-optimised weighting; Fig. 1a shows the trust response using just X_{comms}, with a reasonable deviation in the highlighted node (n_m/Alfa), however in Fig. 1b, where X_{phys} is also included, this deviation is made extremely clear.

Table II summarises the average ΔT_{im} (specifically looking at the malicious node) and shows that on average, including all metrics improves the induced trust variation. It is useful to note however that it is not always "better" to include all metrics; X_{phys} is better at differentiating SlowCoach behaviours, and X_{comms} is slightly better at differentiating MPC and STS behaviours. Some behaviours appear to be difficult to detect in any domain; Selfish Target Selection for example, with it's ΔT_{ix} making it significantly difficult to distinguish from environmental perturbations impacting trust assessment.

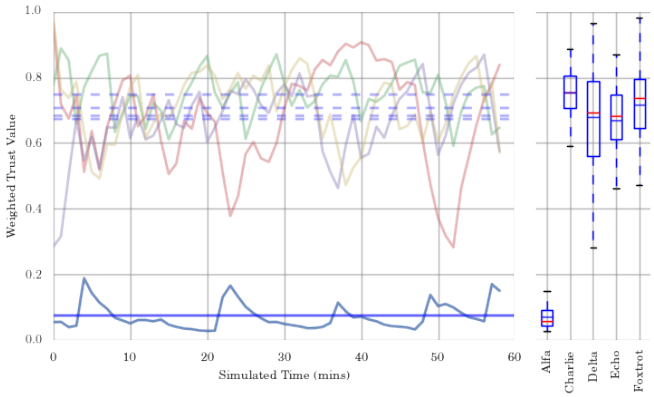
V. CONCLUSION

We have demonstrated that using metrics from across physical and communications domains to assess trust in UANs, this additional information can meaningfully improve induced "deviation" of misbehaviours trust responses to the point where they are more easily detectable and identifiable. We have also demonstrated a purely analytical method for generating optimised metric weighting for MTFM using known behaviour training and Random Forest Regression. In this case, we have naively included all possible metrics, and "arbitrarily" grouped those metrics into the domains they came from, but in interdependent systems such as UANs, these metrics may have significant cross-correlations and redundancies (for instance, Delay could be considered a function of node spacing which is also captured in INDD, which would also impact the received signal strength attained). It may be the case that there exist more performant subsets of X_{merge} that are significantly better at highlighting particular misbehaviours in isolation, and future work will investigate this very case; optimising subset-selection as we have optimised metric-weighting.

Further investigation is required to take this "deviance" and generate a dynamic run-time classifier based on this work. Particularly challenging will be the inclusion of collaborating malicious nodes and periodic misbehaviours; the current construction of ΔT_{ix} assumes that other nodes are fairly reporting their trust assessments, however if several nodes can sufficiently "lie" about how much they trust each other, the cohort based nature of MTFM could be exposed as a weakness.



(a) Comms. Metric Trust Response



(b) Full Metric Trust Response

Fig. 1. Trust Responses for SlowCoach using domain-optimised weights

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] 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

[2] 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

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

[4] A. Bolster and A. Marshall, "Physical Behaviours for Trust Assessmeng in Autonomous Underwater MANETs," in *13th IEEE Int. Conf. Mob. Ad hoc Sens. Syst.* [Online]. Available: http://bolster.online/{~}bolster/16{_}MASS{_}bolster{_}preprint.pdf

[5] L. Breiman, "Random forests," *Mach. Learn.*, pp. 5–32, 2001. [Online]. Available: http://link.springer.com/article/10.1023/A:1010933404324

[6] 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*, pp. 1–10, 2005.

[7] 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

[8] 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*, pp. 142–149, 2011. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6120813

[9] F. Zuo, "Determining Method for Grey Relational Distinguished Coefficient," *SIGICE Bull.*, vol. 20, no. 3, pp. 22–28, jan 1995. [Online]. Available: http://doi.acm.org/10.1145/202081.202086

[10] 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.*, vol. 2, 2010.

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

[12] 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

[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] 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

[15] 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

[16] 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

[17] 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.

[18] 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/