PAP Report 2015
An Investigation into Trust and Reputation Frameworks for Collaborative Teams of Autonomous Underwater Vehicles

Andrew Bolster,
Supervised by
Prof. Alan Marshall and Prof. Simon Maskell

May 19, 2015

1 Overview

- **Started:** October 2011 @ Queen's University Belfast

- **Transferred:** October 2013 @ University of Liverpool

- **Target Submission:** November 2015

The project is focused developing a multi-vector trust management framework (TMF) for collaborative operation of autonomous systems.

Specifically, this project is looking at the relationships between physical behaviour and communications behaviour within teams of autonomous underwater vehicles (AUVs) for uses related to mine counter measures and port protection for defence, as well as persistent survey behaviour for environmental and petrochemical applications.

This work is undertaken as part of The UK-France joint PhD Programme which is jointly managed by Direction Générale de l'Armement (DGA) and Defence Science and Technology Laboratory. It was agreed at the 2010 Anglo-French Summit as one of the ten priorities in 2011 for the Anglo French Defence Research Group (AFDRG). In 2011, five PhDs were funded (two from the UK and three from France), In 2012, the programme grew to nine PhDs (five from the UK and four from France). In 2013 a further ten PhDs were funded (five from the UK and five from France). These PhDs are currently investigating a variety of topics including meta-materials, synthetic biology, sensors, vehicle armour, and human and social sciences as well as this project's work in autonomous underwater vehicles.

2 Summary of Current Outputs

Since the launch of the project, the majority of time has been spent developing a bespoke simulation framework based on Python, developing a variety of "normal", "abnormal" and "malicious" physical behaviours, as well as developing a range of analytical techniques to detect and identify these behaviours to an extremely high degree of statistical accuracy, i.e.in a Monte Carlo style random

repeat experiment, the classifier is correct and confident a significant percentage of experimental runs.

Major outputs and research interactions of the project are:

- Attendance at UComms 2012 (Sestri Levante, Italy)

- Poster Presentations in 2012 (Kassam, Oxford) and 2013 (Heathrow, London)

- Summer Research Placement with DSTL (Software Systems and Dependability for Autonomous Teams)(2013, Portsdown West, Portsmouth)

- Short Paper Presentation to the Association for the Advancement of Artificial Intelligence (AAAI) on "A Multi Vector Trust Framework for Autonomous Systems" (2014, Stanford, CA)

- Technical Report for the UK/US/CAN/AUS/NZ Technical Cooperation Programme ((2014). Analysis of Trust Interfaces in Autonomous and Semi-Autonomous Collaborative MHPC Operations. The Technical Cooperation Program, Technical Report TR-C3I-06-2014) (June 13 - April 14)

- DSTL CDE Collaboration with NPL and Plextek Ltd. on "Precision Timing and Navigation, Challenge 1: Resilient Time and Location Estimation for Networked Assets" (CDE 33135) (Oct 13-May 14)

3   Field Background

3.1   What is 'Trust'?

"Trust" is a word that gets used a lot in many different ways. Mirriam Webster's Dictionary defines trust as "assured reliance on the character, ability, strength, or truth of someone or something".

This rather far reaching definition is very attractive to distributed network design as this "trustworthiness" can be used to inform autonomous actors, or nodes, to the "best" courses and paths to action, making them more efficient and resilient in their operation. Within this context, we define trust as "The Expectation of an actor performing a certain task or range of tasks within a certain confidence or probability".

In the real world use-case of deployable autonomous systems for survey or other application, this trust can take on two real forms; Design Trust, where there is an expectation that a system of systems will perform as specified or designed in operation, and Operational Trust, that is that the individual systems within a larger system will and are performing as designed in the field. It is this area with which we are particularly concerned.

3.2   Trust Management Frameworks (TMFs)

This desire for operational trust in mobile ad hoc networks has lead to the development of several Trust Management Frameworks (TMFs), where such frameworks provide information regarding the estimated future states and operations of nodes within such a network. As such, the operation of such frameworks

has been summarised as "collecting the information necessary to establish a trust relationship and dynamically monitoring and adjusting the existing trust relationship" [3].

Almost all of the work currently applied to establishing trust in mobile ad hoc networks (MANETs) relies either on shared key exchanges with a centralised trust repository (PKI) or with exclusive assessment based on communications behaviour alone, and in that case the vast majority only measure one value (eg packet forward likelihood). Within MANETs, the requirement for distributed trust comes from the decentralised and dynamic communication paths used; if a node moves or an environment changes, the network topology can completely change with paths containing potentially malicious nodes that can disrupt, modify, or reject communications coming from or going to a given node. The motivation for that application of TMFs to MANETs is that by taking historical performance information into accounts, malicious or inefficient actors can be at least detected and routed isolated, preventing further compromise to the distributed network.

These single metric TMFs provide malicious actors with a significant advantage if their activity is undetectable by that one assessed metric, especially if the attacker knows the metric in advance. The objective of operating a TMF is to increase the confidence in, and efficiency of, a system by reducing the amount of undetectable negative operations an attacker can perform. This space of potential attacks can be described as the "Threat Surface". In the case where the attacker can subvert the TMF, the metric under assessment by that TMF does not cover the threat mounted by the attacker. In turn, this causes a superlinearly negative effect in the efficiency of the network.

The TMF is assumed to have reduced the threat surface when in fact it has simply made it more advantageous to attack a different part of it. [2] also raised the need for a more expanded view of trust but did so with a domain-partitioning approach rather than combining trust assessments from multiple domains within networks.

3.3 Multi-Parametric Trust Management Frameworks

Guo developed a form of vectorised trust that allowed the combination and cross-correlation of a range of communications observations (Packet error rate, Noise characteristics, Delay, Throughput, etc) not only as a combined singular value in the form of a Grey Interval, but also presented the ability to use this vectorised trust assessment to not just detect but also identify malicious or abnormal behaviours through weight assessment perturbations [1].

Guo[1] demonstrated the ability of Grey Relational Analysis (GRA)[7] to normalise and combine disparate traits of a communications link such as instantaneous throughput, received signal strength, etc. into a Grey Relational Coefficient, or a "trust vector".

In the case of the terrestrial communications network used in [1], 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,
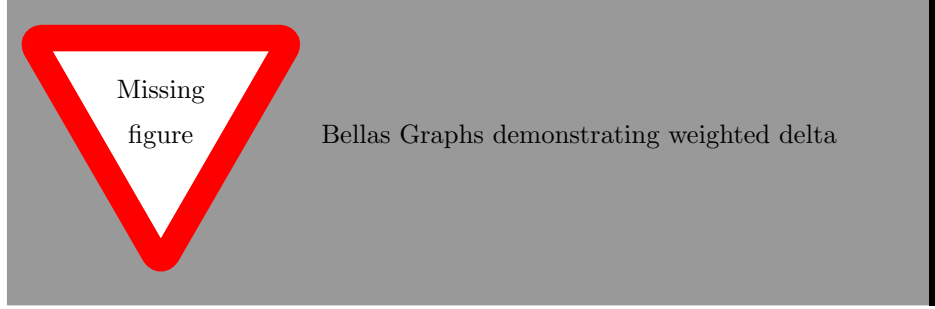
Figure 1: Example of weight-based attack classification taken from (Guo, Marshall & Zhou 2011)

signal strength, data rate, delay, throughput]. The trust vector is given as

$$
\theta_{k,j}^t = \frac{\min_k |a_{k,j}^t - g_j^t| + \rho \max_k |a_{k,j}^t - g_j^t|}{|a_{k,j}^t - g_j^t| + \rho \max_k |a_{k,j}^t - g_j^t|}
$$
$$
\phi_{k,j}^t = \frac{\min_k |a_{k,j}^t - b_j^t| + \rho \max_k |a_{k,j}^t - b_j^t|}{|a_{k,j}^t - b_j^t| + \rho \max_k |a_{k,j}^t - b_j^t|}
\tag{1}
$$

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

For applications involving low fidelity, temporally sparse metrics with unknown statistical distributions, this Grey Relational Analysis is a more stable comparative analysis, providing an interval of potential trust values rather than fuzzy-logic or the Bayesian-Beta distributions found in current TMFs [6].

However, this still ignores the usefulness of physical behaviours in the development of trust assessments, which became the basis for this project.

4 Project Objectives and Novelty

The aims of this project are to extend this "vectorised" methodology to not only combine individual trust metrics (observations) into trust vectors, but to combine such vectorised assessments of trust across "domains" (i.e. combining trust assessments based on observations in the communications domain with trust assessments based on observations in the physical domain).

In single metric trust, such as the use of packet loss rate in other TMFs, this only detects a relatively small area of the potential attack space. Combining several domain specific metrics, in this case in communication, provides not only the capability to detect a much wider range of attack types, but also to discriminate between them, potentially disclosing of the tactics of an attacker.

Our hope is that by using multiple domains together, we can provide a higher level, strategic point of view on an attacker or attackers, enabling the generation

of preventative and reactive strategies to defend against them. This would leave even the most knowledgeable attacker with no option but to behave properly to avoid detections, eliminating any potential reward they could achieve without detection.
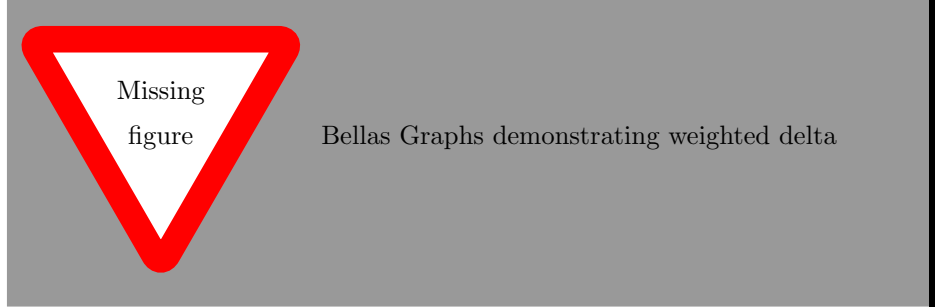

Figure 2: With the addition of further metrics, the threat surface available to an attacker is greatly reduced

Our main goal in this project has been to investigate if the methodologies that Guo applied to communications, can equally be applied to physical behaviour. Our approach to this has been to develop an agent based simulation platform built in Python, emulating both the physical and communicative environments for teams of flocking nodes performing some task, such as mine counter measure survey, port protection, mother-ship protection and others.

This interaction in the physical world opens up a range of metrics that can be used for trust assessment, in an effort to further restrict undetected malicious behaviour on the exposed threat surface of a network.

5   Current Progress and Results

5.1   Physical Behaviours

Through collaboration NATO's Centre for Maritime Research and Experimentation based in Italy, and the UK's Defence Science and Technology Laboratory, we settled on a range of observable metrics based on the position and attitudes of other nodes in the group. These were;

- The Inter Node Heading Deviation (INHD), i.e. the deviation in heading from a local group average,

- The Inter Node Distance Deviation (INDD), i.e. the deviation of a given node's position with respect to the average node spacing across the rest of the local group,

- And The Node's Absolute Speed.

Along with metric selection, we also arrived at a few sample 'Misbehaviours' covering both malicious and non-malicious non-optimal behaviours which were simulated in the framework. These were:

- **The Shadow**, where a node is following the fleet without appropriate mission knowledge such as the waypoints in a patrol path, modeling an 'infected' or masquerading node in the fleet

- **The Spy**, where a node consistently or intermittently rises to the top of the fleet, potentially surfacing to relay mission information to an unauthorised third party via a back-channel communications channel such as RF

- **The Sloth**, where a node is selfishly conserving energy by not making complete patrol paths or consistently taking a minimal path around the tolerances of the waypoint path

- **The Stalker**, where a node in a multi-node system preferentially 'tails' a given node above others

- **The Scoundrel**, where a node falsely reports it's position or velocity with the intention to corrupt any collaborative positioning system

- **The Slow Coach**, where a node is operating correctly but has a defect in it's power train causing reduced maneuverability, i.e. a runtime defect in an otherwise good node

- **The Spin Doctor**, likewise is operating correctly but it's Inertial Navigation System is damaged and consistently bears left/right by some level.

Initially we've only directly considered Shadow and Slow Coach behaviours, as this presents a useful side effect of trust based on physical behaviour; the ability for a network to self-test it's health by discerning the difference between broken and malicious behaviour.

In a simple port protection scenario, where nodes are patrolling around a series of waypoints around an area, We can demonstrate this detective and selective power through a series of graphs that clearly justify the abstraction of such metrics into a trust value.

Figure 3: Charts showing deviation activation curves for three different simulated behaviours, with a fuzed trust value in the bottom row.

Each vertical of this chart shows a different behaviour; with the baseline waypointing behaviour in the middle, flanked on the left by the malicious "Shadow"

behaviour and on the right by the benign but sub-optimal "Slowcoach" behaviour. The per-node metrics as described are shown in the horizontal, with Internode Heading Deviation at the top, followed by Absolute Speed and Distance Deviation in the middle and finally a combined trust 'vectorised' (dis)trust assessment in the bottom, based on a decaying windowed average of instantaneous trust assessments across observation types.

With this chart it's shown that the Slowcoach behaviour manifests itself in the Speed metric. more so than the Shadow behaviour. Using this information we can build a trust weight vector to discriminate between these two ostensibly similar behaviours.

With our current analysis we can do so with an average 96% positive identification rate with a 1% false positive rate, i.e. instances where a Slowcoach is confidently detected as a Shadow and visa versa. (It should be noted that in this analysis, there were no cases where the malicious behaviour (Shadow) was misdetected as the purely abnormal/faulty behaviour (Slow Coach).

6 Changes to Assumptions

Over the course of the project, initial assumptions made about the nature, operation, and context of the project have changed wildly. This is largely due to the increased access to current and planned operational data from the Royal Navy while at DSTL Portsdown West, working with both the Naval Systems and Information Systems teams specifically on the area of maritime autonomy and the use of increasingly decentralised and autonomous systems for marine survey and observation. This work also contributed to a transnational collaboration looking at the establishment of systematic trust in autonomous systems.

Likewise, our work with Plextek and NPL on collaborative positioning using high precision timing was extremely helpful in developing my awareness not just of positioning techniques used by current platforms but also increased knowledge and time to generate simulation extensions to more accurately model not only the errors in these positioning systems but also algorithms for estimating the propagation time of acoustic comms.

Initial Assumption	Correction
AUV operation in open water	AUVs (currently) mainly operate in coastal/littoral/riverbed areas with highly dynamic SSP (Speed of Sound Profile) and environmental variations. Shallow environmental profile
AUV sensing dynamically adjusts to the environment and motion	SSS (Side Scanning Sonar) is extremely sensitive to not only height but strafe drift and can corrupt entire missions by being out of alignment more than 8-10m horizontally and 2-4m vertically (depending on initial configuration) SAR (Synthetic Aperture Sonar), which builds up acoustic profiles over several "pings" is slightly more tolerant but any variation in path massively increases computational complexity of imagery (Possibly a "Moore's Law" Problem)

Initial Assumption	Correction
AUV positioning is equally bad in all dimensions due to loss of GNSS, relying on INS, with cumulative errors on the order of tens of metres per hour of operation	Depth sensing is extremely accurate INS errors are directionally cumulative based on a electro/gyroscopic bias in the device. This means that if you travel two hours in one direction and two hours back in the opposite direction, your cumulative INS (Gyroscopic) error will be minimal Additionally, the use of bottom tracking DVL (Doppler Velocity Log) provides quite good instantaneous velocity (through water) tracking but with normal errors. These errors are relative to the height above floor and cumulative with time. All in all for most cases, drift can be characterised at around half the expected rate (5/hr) Simulation Model Characterisations have been generated for both Gyro and DVL error statistics Further, collaborative measures can be taken to normalise these errors and consistently improve position accuracy by up to 40%. This collaborative system has also been implemented into the simulation framework, and directly leads into the Scoundrel malicious behaviour model.
— Nodes operate with a spherical distribution (i.e. vertically and horizontally distributed)	All of these together lead to the the improved assumption that nodes operate in a flat arrangement, maintaining a known altitude for tracking accuracy, DVL consistency, SSS/SAR resolution and SSP stability.
Acoustic Comms is terrible and curved	This assumption was validated and expanded upon, with the generation of a simplified "Bellhop" simulation model generated to reasonably estimate the time of flight characteristics (but not ISI) of acoustic comms using ray tracing. Current bitrates for ¡500m ranges are around the 100kbps scale, providing an upper limit on comms bandwidth for inter-node collaboration
AUV operations will be mostly isolated	Currently this is not the case, with AUV MHPC operations mainly based on either shore-based teams or Hunt/Sandown class based surface vessels in close proximity to operational area. However BLOS (beyond line of sight) AUV operation is part of the MoD's Future Force 2020 programme so it is reasonable to continue operating as if that is the case.
AUVs will be "weaponless"	While there are direct-disposal tethered ROVs fitted with explosive payloads, this assumption has largely borne out; the biggest blockade to having weaponised AUVs is legal trust in the security of operation of such an autonomous system, making it extremely unlikely that decentralised AUVs will ever be legally weaponised.

7 Development, Publication Plans

7.1 Behaviour Detection

Current work is concentrated on the improvement of the analysis of behaviours, with the intention of submission for publication of a paper on the threat detection methodology over the summer conference season. Potential conferences suitable for this publication are: (Submission deadlines in parenthesis)

ASON 2014 : Seventh International Workshop on Autonomous Self-Organizing Networks (Aug 14)

AHUC 2014 : The Fourth International Workshop on Ad Hoc and Ubiquitous Computing (Aug 14)

ICCAR 2015 : WASET International Conference on Control, Automation and Robotics (Dec 14)

7.2 MANET/Marine comparison

Est: Sept 14

In parallel with completing the analysis of the behaviour and the behaviour classifier, a more formal comparison of the differences between the classical terrestrial MANET space and the Marine operating environment will be undertaken.

7.3 Multi-Domain Trust Assessment

Est: Nov 14

Following on from this work, it is intended to focus on the cross domain implementation of trust, which is, combining communicative and behavioural trust assessments.

This body of work has two main areas; analytical and critical.

Initially, It is expected that this work will be largely analytical, maintaining context-free analysis for multi-domain trust but utilising AUV MANETs as an exemplar implementation for critical assessment of proposed cross-domain combination strategies.

This will require the completion of the implementation of existing communications TMF's for comparative assessment, including but not limited to Beta Expectation Values [5], Objective Trust Management Framework[4], and Multi Parameter Trust Framework [1]

Ideally this work will go on as a Journal paper into potentially IEEE Transactions on Communications, Dependable and Secure Computing, or Intelligent Systems

7.4 Reactionary/Perturbative Trust

Est: Feb 14

After the development and analysis of multi-vector trust, it is intended that some time be spent investigating reactionary behaviours, both communicative and physical, to allow a distributed, sparsely connected, decentralised team of nodes to dynamically "interrogate" or "test" the trustworthiness of nodes within the team.

Aim is for conference submission to SASO15:Self-Adaptive and Self-Organizing Systems, or SEAMS15: Software Engineering for Adaptive and Self-Managing Systems (both unconfirmed)

8   Proposed Thesis Chapter Titles

- Abstract, Acknowledgements, Introduction,

- Background Information on Trust and it's applications to MANETs

    - Discussion on abstract analysis of trust networks
    - Discussion on the threat surface of Mobile Ad Hoc Networks and how that has been protected so far
    - Introduction to Trust Management Frameworks and their benefits

- Background Information on Maritime Uses of Autonomous Systems

    - Discussion of current and future approaches to areas where autonomous systems can be used mainly focused on Mine counter measures, Hydrography and Patrol Capabilities (MHPC)
    - Discussion of the contextual human factors around integrating autonomous systems into existing human-based solutions.

- Trust in Autonomous Systems of Systems for Maritime Defence Applications

    - Predominantly following on from work already accomplished under ?Analysis of Trust Interfaces in Autonomous and Semi-Autonomous Collaborative MHPC Operations?, including development of representative malicious and abnormal behaviours

- Strategies for Multi-Domain Trust Assessment

    - Analytical establishment of Multi-Domain Trust, from an information theoretic standpoint.

- Modelling and Analysis of Collaborative Node Kinematic Behaviours in Underwater Acoustic MANETS

    - Touching on the development of the simulation platform but focused on the mobility and assessment of mobility between nodes, including identification of suitable motive metrics and analyses of these motions to establish intent or abnormality
    - Passing mention of work done in Drift analysis with NPL/Plextek as supporting evidence

- Comparative Analysis of Multi-Domain Trust Assessment in Collaborative Mobile Networks

- Investigation into the relative performance characteristics of multi-domain combination strategies in an exemplary context (AUV teams) against existing single and multi metric TMFs

- Reactionary Behaviours to increase decentralised trust in isolated environments

    - How can nodes interrogate each other, both communicatively and physically, to test or establish trust beyond the reach of an agreed authority

- Conclusions, Bibliography

References

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

[2] Dijiang Huang, Xiaoyan Hong, and M Gerla, *Situation-aware trust architecture for vehicular networks*, Commun. Mag. IEEE (2010), no. November, 128–135.

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

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

[5] Ruidong Li Ruidong Li, Jie Li Jie Li, Peng Liu Peng Liu, and Hsiao-Hwa Chen Hsiao-Hwa Chen, *An Objective Trust Management Framework for Mobile Ad Hoc Networks*, 2007 IEEE 65th Veh. Technol. Conf. - VTC2007-Spring (2007), 56–60.

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

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