

Towards Distributed, Behaviour-Based Trust for Autonomous Submarine Teams

Andrew Bolster, Prof. Alan Marshall, Prof. Jean-Guy Fontaine Advanced Networks Research Group, University of Liverpool, UK



Project Background

- ► Project launched at QUB ECIT in 2011 under the DSTL/DGA Anglo French Defence Research Group PhD Programme
- ► What lessons from the Mobile Ad Hoc Network (MANET) space can be transferred to the marine environment?
- ► Teams of 3 16 Autonomous Underwater Vehicles (AUVs) with Mine countermeasures, Hydrography, and Patrol Capabilities (MHPC)
- ▶ Defence focus, assumption of highly capable enemy attempting to compromise communications / operations
- attempting to compromise communications / operations▶ Primary Simulation/Analysis work done in 12/13
- ► Moved to UoL Oct 13 after 2 mth placement @ DSTL PDW Naval Systems / Information Systems departments.
- ► CDE Project on Precision Timing for Collaborative Acoustic Positioning with NPL/Plextek

Introduction

Aim of project: To use physical behaviours and observations to assess and maintain trust within mobile, marine, ad-hoc networks

Small fleets of AUVs (*Autonomous Underwater Vehicles*) will be expected to operate in isolated environments.

This requires an auditable sense of trust within the remote intra-fleet communications networks, incorporating

- ► Communications Activity
- Mission Suitability/Capability
- Behavioural Monitoring

The use of centrally coordinated trust models presents a single point of failure.

Secure communication in marine environments is expensive and time consuming; adopting a decentralised form of trust assurance will reduce these costs by localising the per-node security environment.



Figure 1: REMUS 100 AUV, as deployed at CMRE, a potential target platform for this work

Trust

Trust is:

- ► The expectation of an actor performing a certain task or range of tasks within a certain confidence or probability
- a belief on the reliability of an entity
- ► based on both direct and indirect historical experience

Individual trust opinions are shared within the network concerning a range of activities:

- Transmission Relaying (Local and/or Backhaul)
- Position Relaying
- Reporting Accuracy

These Trust opinions also apply to extra-fleet entities, such as surface platforms, submarine comms links, and coastal stations, allowing the fleet to collaboratively form an opinion of these actors.

Trust Management Frameworks (TMFs)

TMFs are protocols designed to provide information regarding the estimated future states and operations of nodes within networks

"[...]collecting the information necessary to establish a trust relationship and dynamically monitoring and adjusting the existing trust relationship" - [1]

Enables nodes to form collaborative *opinions* on their cohort nodes based on

- Direct Observation of Communications Behaviour (eg Successfully Forwarded Packets)
- ► Common-Neighbour Recommendation
- ► Indirect Reputation

Multiple transitive relationships can be maintained over time, providing trust resilience with dynamic network topologies

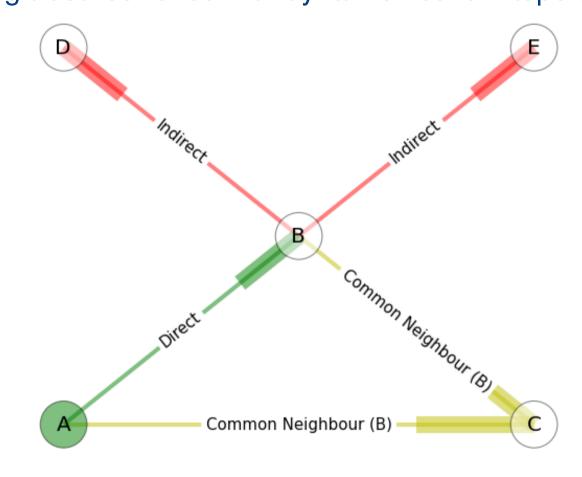


Figure 2: Direct, Recommendation, and Indirect trust relationships

The Need for Multi-Domain Trust in Autonomous Systems

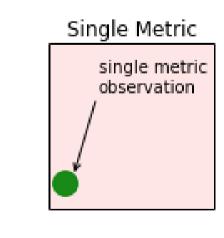
Communications not the only target for an attacker (or failure);

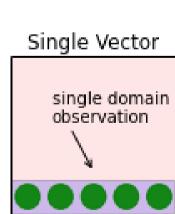
- ► Following to restricted area
- Masquerading
- ► Hardware Degradation
- ► Resource attack via propulsive power

Physical observation presents opportunity to further reduce the available threat surface while also discriminating between 'True' attacks and mechanical failure.

Also could provide additional 'handshake' protocols for 'friendly' fleets/teams through reactionary behaviours

Potential attacks exist within a multi-domain threat surface, and as further metrics and domains of trust are included in a TMF, attackers are increasingly restricted in their behaviour until the only way to avoid detection is to behave correctly.





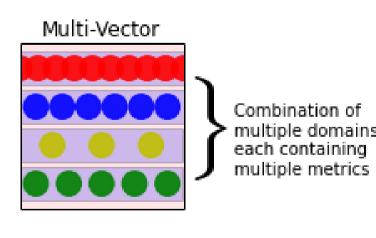


Figure 3: Threat Surface for Trust Management Frameworks

Operational Mission Profiles

Generic behaviours currently under investigation include:

- Waypointing Attraction to a point or a chain of points, providing pre-described patrol networks
- Surveying Fleets can be tasked to provide one-shot, or persistent coverage of an area of the environment utilising a dynamic lawnmower pattern.
- ► Dynamic Constraint Repulsion from a series of points, analogous to sea-borders or shipping lanes.
- Potentially Exploitable Behaviours not yet developed include:
- ► Capacity Based Homing Where a node leaves and later returns to the fleet, for instance for refuelling or resupply.
- ► Dynamic Communications Maintenance The fleet can adjust to changing communications environments.

Malicious Behaviours

A series of 'Malicious' behaviours have been designed, where one or more nodes were actively attempting to compromise the fleet in some way.

- ► Shadow Following the fleet without restricted mission data
- ► Spy A node that intermittently rises in the fleet, potentially surfacing to relay mission information to a third party
- ► Sloth Selfish conservation of energy by taking minimal paths
- ➤ Stalker Tailing a specific node, attempting to use this consistent history to poision the trust network
- ➤ Scoundrel Falsely reporting its estimated position with the intention to corrupt a collaborative positioning system or induce a collision

Additionally, there are non-malicious behaviours, the classification of which is equally important in separating friend from foe.

- ► Slow Coach Where a node's power train is damaged, causing reduced maneuverability and performance
- ➤ Spin Doctor Damage to control surfaces or to Inertial Navigation System

Proof of Concept Analysis

The aim of the Proof of Concept is to demostrate not only the detection capability of using behavioural metrics to access trust but to differentiate between malicious and impaired bheaviours.

Three fleets are simluated performing a simple patrol mission, each with a particular behaviour set. The first has a malicious node performing a 'Shadow' behaviour, representing a malicious outside actor attempting to infiltrate the fleet. The second has an impaired node exhibiding 'SlowCoach' behaviour, representing a damaged or faulty but otherwise good node.

The simulations were run to for a standard eight hour mission time, with three primary metrics being assessed at run time by each node:

- ► *INHD*: Inter Node Heading Deviation, or the per-node deviation from the fleet-average velocity vector
- ► Node Speed: The Magnitude of Velocity of each node
- ► *INDD*: Inter Node Distance Deviation, or the normalised deviation between the inter-node distance matrices from the perspective of each node

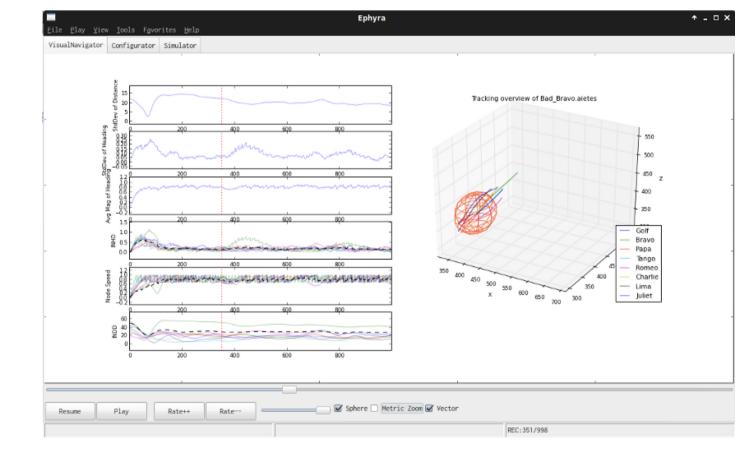


Figure 4: The Ephyra visualiser allows for rapid modelling and analysis of node data, as well as overlay information on individual and fleet behaviours and metrics.

In this model, there are no 'positive' trust demonstrations, so all nodes are 'distrusted' to some degree or 'untrusted'. This distrust is dependant on consistent or periodic deviation from fleet norms. This trust assessment is developed by each individual node against each other node, however for this Proof of Concept, a global metric state was used (free of environmental and positional drift factors).

Analysis Cont.

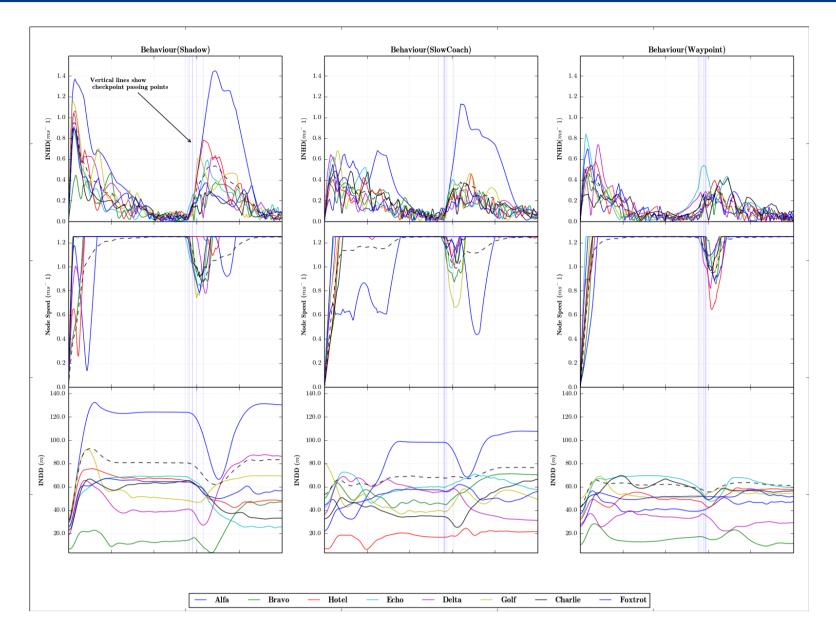


Figure 5: Per-Simulation metric measurements for Malicious, Impaired, and Control behaviours (Shadow, SlowCoach and Waypoint)

INDD is an obvious candidate for a suspicion 'trigger', but looking at *INHD* values after a few hundred seconds of simulation time; an anomaly is clearly being detected.

In addition, *Alfa* node (the Blue Line) is clearly an outlier in terms of *INHD* and *Node Speed* in the earliest sections of the graph, at which point nothing appears to be out of the ordinary in *INDD*. This implies that a fusion of metrics would be more effective than a simple detection envelope on a single metric.

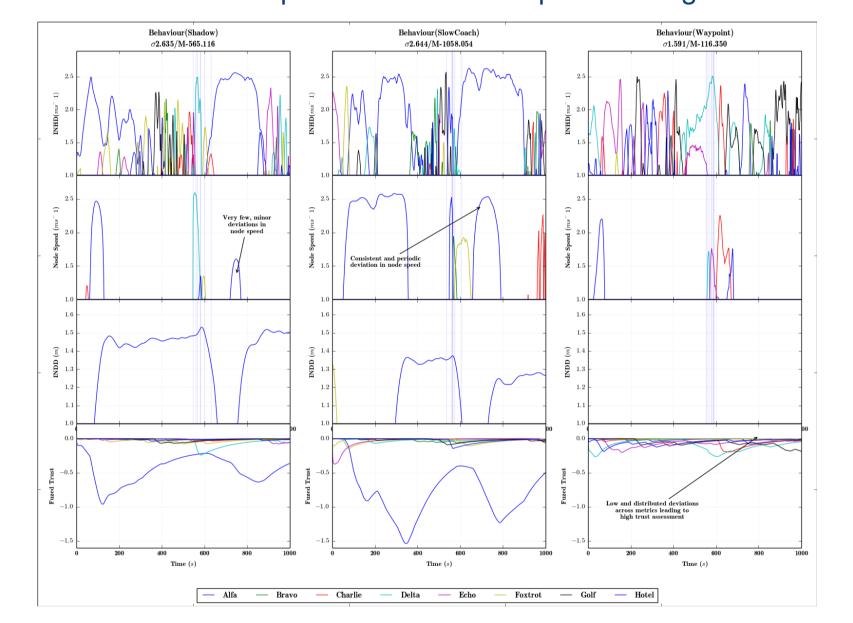


Figure 6: Per-Node deviations for each metric, with an additional row showing an EWMA based cross-metric trust assessment. Note the different in 'Node Speed' triggers between the malicious and impaired behaviours

Considering the Baseline data (Right side of Figure 5), it's clear that these metrics are not infallible, as is demonstrated by the number of relatively short-lived false positives, demonstrating the need to use multiple metrics for reliable trust assessment.

Figure 6 demonstrates a windowed, weighted trust fusion, where deviations in individual metrics are combined to generate a Trust Value.

Future Applications

- ▶ Due to the high communications, motion, and computation costs, and lack of external location reporting (e.g. GPS), behavioural analysis in the marine environment is particularly difficult, but if successful, can be reliably applied in a wide variety of fields including but not limited to
- Self-Driving Cars
 Self-Driving
- ▷ Environmental Survey drones (terrestrial, marine, and aerial)
- ▶ Internet Certificate Authority verification
 ▶ Verificate Distributed Computing
- Verifiable Distributed Computing

Conclusions

This research area presents a range of challenges and opportunities within both civil and defence operations; an auditable trust framework for automated marine craft would be a significant enabling factor to the roll-out of more low-maintenance or even "Fire and Forget" deployments for persistent patrol/monitoring tasks.

Open Hypotheses in this field that this project intends to answer are:

- ► How can optimality in trust assessment based on behaviour be defined win a distributed, dynamic network topology?
- ► Is there a quantifiable benefit to cross-domain comparison beyond single-vector trust? (i.e. 1-D vector vs cross domain comparison)
- ► Is there an optimal *generic* cross domain fusion methodology?

Current Publications

- ► A Multi-Vector Trust Framework for Autonomous Systems [2]
- ▷ Symposium paper to the Association for the Advancement of Artificial Intelligence on the current state of work, presenting our progress towards multi-vector trust
- ► Analysis of Trust Interfaces in Autonomous and Semi-Autonomous Collaborative MHPC Operations [3]
- Part of a Five-Eyes defence strategy programme (TTCP) for assuring C3I capabilities as part of FF2020

Development Plan

- ▶ Behaviour Detection (Q3 14) Formal Analysis of Behavioura Trust Systems
- ▷ ASON 2014 : Seventh Int. WS on Autonomous Self-Organizing Networks (Aug 14)
- ▷ AHUC 2014 : The Fourth Int. WS on Ad Hoc and Ubiquitous Computing (Aug 14)
- ▷ ICCAR 2015 : WASET Int. Conf. on Control, Automation and Robotics (Dec 14)
- ► MANET/Marine comparison (Q4 14) Formal Comparison between Terrestrial MANET / Marine contexts
- Multi-Domain Trust Assessment (Q4 14) Combination of Communicative and Physical Behaviour Trusts
 IEEE Trans. on Communications / Dependable and Secure
- Computing / Intelligent Systems
 ▶ Reactionary/Perturbative Trust (Q1 15) Exploration of reactionary behaviours for teams to 'shake down' suspects
- SASO15:Self-Adaptive and Self-Organizing Systems
- SASO 15:Sell-Adaptive and Sell-Organizing System
 SEAMS15: Software Engineering for Adaptive and Self-Managing Systems

Bibliography

- Huaizhi Li and Mukesh Singhal. "Trust Management in Distributed Systems". In: *Computer* 40.2 (2007), pp. 45–53. ISSN: 00189162. DOI: 10.1109/MC.2007.76. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4085622.
- Andrew Bolster and Alan Marshall. "A Multi-Vector Trust Framework for Autonomous Systems". In: 2014 AAAI Spring Symposium Series. Stanford, CA, 2014, pp. 17–19. URL: http://www.aaai.org/ocs/index.php/SSS/SSS14/paper/viewFile/7697/7724.
- Andrew Bolster. *Analysis of Trust Interfaces in Autonomous and Semi-Autonomous Collaborative MHPC Operations*. Tech. rep. The Technical Cooperation Program, 2014.