

TAS-S FINAL REPORT OCTOBER , 2024

UKRI AUTONOMOUS SYSTEMS NODE IN
SECURITY



Foreword

The UKRI Trustworthy Autonomous Systems Node in Security (TAS-S) constitutes an exciting opportunity for collaborative research that is additionally supported by our unique security and Autonomous Systems (AS) test beds.

The project has extensive stakeholder support, both domestic and international, from academics to AS providers, AS users and AS regulators.

Further details on all of the sections in this report can be found on the TAS-S website [TAS-S website](#).

This report gives an update on our Node's activities since the publication of the Mid-Year report in August 2023.

We have had a busy and exciting second half of 2023 and first half of 2024, with an impressive number of publications and a wide range of engagement activities.

We are looking forward to continuing to work with internal and external colleagues, and to the publication of our Final Report this year.

Professor Corinne May-Chahal,
Principal Investigator

Contents

01	Introduction	p.3
02	TAS-S Team	p.4
03	Research overview	p.6
04	RS1 Activity Summary 2020-2024	p.12
05	RS2 Activity Summary 2020-2024	p.21
06	RS3 Activity Summary 2020-2024	p.36
07	Case Studies	p.45
08	Publications, Media & Products	p.53
09	Engagement Activities	p.65
10	Acknowledgements and contact details	p.68

1. Introduction

Autonomous Systems (AS) can be broadly categorised as the ability to effectively conduct a mission with varied levels of “absence of human intervention” including completely unsupervised operations. Typical examples, spanning an ever-growing diversity of civilian, industrial and military applications across terrestrial, aerial and aquatic environments include autonomous vehicles, industrial automation, assisted living and a variety of logistical support to complement and supplement societal needs.

As technologically complex and networked cyber-physical entities, an AS needs to ensure “safe and secure” mission functionality despite the occurrence of any encountered cyberphysical disruptions. As such, an AS is a highly-dynamic entity that needs to adapt to the vagaries of its operational environments and security profiles (including changing threats). Providing “predictable, scalable and composable” security (of the AS assets, of the AS operations and the AS usage environment) in “uncontrolled and dynamic” operational environments is the objective of TAS-S.

The TAS Security Node’s research is centred around a seamless collaboration between fundamental cross-disciplinary security research and autonomous systems research at Lancaster and Cranfield Universities. To accomplish this vision, TAS-S utilizes interlinked cross disciplinary Research Strands (RS) to address 3 core challenge areas in autonomous system (AS) security:

*Research Strand 1(RS1):
'Securing the AS "usage" environment'*

*Research Strand 2(RS2):
'Can we secure the AS "operations" environment?'*

*Research Strand 3 (RS3):
'Can we secure the AS "user" environment?'*

2. TAS-S Team 2023-2024

TAS-S assembles a cross-disciplinary team of internationally reputed security experts from Lancaster and Cranfield Universities who are based across a wide range of research areas including Distributed Systems, Controls, AI, Communications, Sociology and Law.

2023-2024 Research Strand Leads and Project Manager



Prof. Neeraj Suri,
PI (11.2020-3.2024),
RS1 Lead
Lancaster University



Prof. Weisi Guo
Co-PI,
RS2 Lead
Cranfield University



Prof. Corinne May-Chahal
PI (4.2024-10.2024),
RS3 Lead
Lancaster University



Tetiana Vestel
Project Manager
Lancaster University

Co-Is, Lancaster University

Prof. Plamen Angelov
Prof. Joe Deville
Prof. Catherine Easton

Co-Is, Cranfield University

Prof. Gokhan Inalhan
Prof. Antonios Tsourdos
Dr. Lisa Dorn

Postdoctoral Researchers, Lancaster University

Dr. Zhengxin Yu
Dr. Andrew Sogokon
Dr. Luke Moffat
Dr. Yi Li
Dr. Yang Lu

Postdoctoral Researchers, Cranfield University

Dr. Zhuangkun Wei
Dr. Yun Tang
Dr. Aykut Cetin
Dr. Emre Saldiran

PhD Students, Lancaster University

Xavier Hickman
Alvaro Lopez
Ovini Gunasekera
Julia Michelin Alvarenga

PhD Students, Cranfield University

Ajinkya Lakhepatil

2. TAS-S Team 2020-2024

TAS-S cross-disciplinary team comprised 34 academics and professionals from more than 10 countries including UK, USA, China, Sweden, Singapore, Turkey, Spain, Brazil, and Ukraine.

PIs, Co-PI & Co-Is

Lancaster University

Prof. Neeraj Suri
Prof. Corinne May-Chahal
Prof. Plamen Angelov
Prof. David Hutchison
Prof. Daniel Prince
Prof. Joe Deville
Dr. Catherine Easton
Dr. Vasileios Giotsas

Cranfield University

Prof. Weisi Guo
Prof. Gokhan Inalhan
Prof. Antonios Tsourdos
Dr. Lisa Dorn

Postdoctoral Researchers

Lancaster University

Dr. Zhengxin Yu
Dr. Andrew Sogokon
Dr. Luke Moffat
Eduardo Almeida Soares
Pierre Ciholas
Dr. Yi Li
Dr. Yang Lu
Dr. Samson Palmer

Cranfield University

Dr. Burak Yuksek
Dr. Zhuangkun Wei
Dr. Oscar Gonzalez Villarreal
Dr. Anders åf Wahlberg
Dr. Yun Tang
Dr. Aykut Cetin
Dr. Emre Saldiran

PhD Students

Lancaster University

Xavier Hickman
Alvaro Lopez
Ovini Gunasekera
Julia Michelin Alvarenga

Cranfield University

Ajinkya Lakhepatil

Project Managers

Lancaster University

Pamela Forster
Tetiana Vestel

3. Research overview

Securing the AS ‘usage’ environment. Establishing the fundamental AS ‘usage’ framework for providing and assessing multilayered, multi-dimensional adaptive AS security in dynamic mixed mode environments. A key innovation was to establish a comprehensive formal specification of the dynamic AS environment as a data flow model, the specification of AS behaviours and the corresponding threat models. This unique hierarchical federated learning framework was subsequently validated in GPS-denied urban AS environments. RS1 also proposed a progression of innovations in detecting and mitigating adversarial AS attacks validated on a runway landing system.

In working towards securing the AS ‘Operations’ environment our objective was to ascertain exposure (and the consequent mitigation) of AS ‘operations’ to cyber-physical attacks by characterizing the attack surfaces (i.e. entry points and likelihoods) across the mission, control and information surfaces in a technology and mission invariant manner. Our innovations include both: (1) red-teaming attacks (generative evasion attacks to manipulating electromagnetic channels using meta-surfaces), and (2) developing entirely new cross-layer defences – using swarm drone control physics to generate network cipher keys. We also began to realise that in realistic AS scenarios, human users in the loop, or in the environment may have trust and security concerns that are difficult to capture numerically in pre-programmed routines. Here, our last advance is to use large-language-models (LLMs) to develop ethical-social value tuned personalised autonomy that is cognisant of both risks and wider human values.

To secure the AS ‘User’ environment we aimed to ascertain human and social threats with a focus on AS interaction with both human behaviour and social context. Autonomous transportation, and specifically National Highways, were selected as a domain that could capture generic threats in the user environment. Methods included systematic review, creative arts-based workshops, a panel survey, focus groups with members of the public, and an experimental demo using LLM to encode ethical reasoning in basic transport decisions. Human understanding of threats derives from a lack of trust in AS, in terms of their lack of testing, response to unpredictable scenarios (such as major weather events), fear of hacking, malfunction, and loss of jobs. Concerted efforts must be applied to meaningful public engagement with AS development if technical advances are to progress successfully. Organisations transforming their operations should focus attention on social values and priorities for achievable and secure AS application.

3. Research overview

Outcomes

(for further information see Activity Summaries and Publications, Media & Products)

RS1A: Development of specifications of threats and functional behaviour in Autonomous Systems – Research: The specification of Autonomous System (AS) threats along with the fundamental problems of formal verification and properties in models of AS embedded in a physical environment (such systems typically combine discrete and continuous behaviour and are thus examples of Cyber-Physical Systems). Outputs: focus on adversarial image threats (1), specifications for autonomous systems (2), Real Arithmetic in TLA+ (3), and synthesis of barrier certificates using linear programming relaxations.

RS1B: Development of Frameworks for Federated ML in Autonomous Systems - Research: the progressive development of a framework for federated ML that can handle the AS-relevant aspects of data heterogeneity, adversarial data/model corruptions and incorporation of mobility aspects on the data. Outputs focus on data heterogeneity (4), adversarial data/model corruptions (5), visual-inertial navigation (6) and mobile AS nodes (7)

RS2: Development of GNSS-Denied Navigation - Research: A robust-by-design Federated Meta Learning based visual odometry algorithm to improve pose estimation accuracy, adapt to environments by using differentiable meta models and tuning its architecture to defence against cyber-attacks (8).

RS2 B-C: Development of Secure Control & Communications – Research: A novel form of physics-driven digital encryption in swarm ASs. Outputs: a new security mechanism called control layer security (9, 10).

RS2 A-B-C: Inferring the Security Risk of an Uncooperative AS - Research: the mission, control and information layer as a novel form of physics-informed inverse learning against uncooperative or suspicious AS. Outputs: A control-physics informed machine learning (CPhy-ML) was developed that can robustly infer across intention classes, achieving a 48.28% performance improvement over traditional methods (11), and an accurate trajectory inference algorithm (12).

RS2 C: Dealing with Sparse, Biased, and Adversarial Training Data for Uncooperative Target Recognition – Research: Understanding the general distribution of the drone data via a generative adversarial network (GAN) and explaining the under-learned data features using topological data analysis (TDA) (13). Evasion patterns for soft-body human stakeholders, we consider this a new adversarial training area (14).

3. Research overview

RS2 C: Identifying and Countering the Risks of Adversarial Meta-Surfaces for Secure Communications – Research: We designed Eve-RIS schemes against two PL-SKG techniques and Man-in-the-middle malicious RIS (MITM-RIS) eavesdropping. Outputs: The proposed Eve-RIS can achieve a high key match rate with legitimate users and is resistant to most of the current defensive approaches, identifying new areas for adversarial research (15). Simulation results GAN-based and symbolic-based PL-SKGs can achieve high key agreement rates between legitimate users and are resistant to MITM-RIS Eve with the knowledge of legitimate feature generation (NNs or formulas). This therefore paves the way to secure wireless communications with untrusted reflective devices in future 6G (16).

RS2 C: Identifying Cybersecurity Risks and Mitigation Measures for Internet of Nano/Bio Swarms – Research: new vectors of attack and new methods of PLS, focusing on 3 areas:

- (1) information theoretical secrecy bounds for molecular communications,
- (2) key-less steering and decentralized key-based PLS methods, and
- (3) new methods of achieving encoding and encryption through bio-molecular compounds. (17, 18).

RS3A Individual Behavioural Adaptation – Research: How trust in AS is built and lost to address the learning and appropriation, behavioural adaptation to ADAS and the safety effect of ESC. Outputs: Trust in regulators, system manufacturers and experience had significant correlations with trust in AS (19), a scale for canvassing experience to measure impact on adaptation (20). ESC safety effects ranged from 38 to 77 percent reduction of crashes but effects differed depending on the methods used (21) and a downward trend in crashes was present before ESC introduction, but trends thereafter were weaker (22).

Ethical, Legal and Social Issues (RS3 B&C) – Research: Evaluation of securitized AS risks and societal response in end-user environments. Incorporating ethics and social values into AS design and understanding how human-machine cognitive interaction constitute and regulate the making of distributed autonomous technology. Outputs: Ethics in AS research (23), relational approaches to AS ethics (24), the ELSI-AS Toolkit (see case study 2), encoding ethics into AS (25) and distributed epistemic systems (26).

3. Research overview

Metrics: Key outputs and associated impacts

- 1. Publications:** 62 publications with 219 citations with a range of 0-21, up to October 3rd, 2024
- 2. Frameworks:** a) Env. for formal specification of AS ops and attacks (w. Airbus, TTTech), b) Env. for GPS & commn-denied navigation (w. BAE Systems) (see case study 1).
- 3. Standards:** TAS-S input, as a BSI expert, on 'ISO TC241 Road Traffic Safety Management Systems WG6 on Ethical Considerations for AV's' incorporated in publication of standard ISO 39003 in August 2023 (see Case Study 3).
- 4. Funding+:** a) 2 RAEng security fellowships in AS (drones, and air traffic management), b) Co-leading UK Future Commn hub in understanding how ASs shape requirements of future 6G network design (Case Study 1).
- 5. Awards/ECR Development:** Best Doctoral Thesis Award, Best Paper Award @MASS'23; ECR's placed at IBM Research, Caterpillar, In-Space Missions...
- 6. Spinout:** AI-CyberSec 'Mindgard' with 3M VC funding
- 7. International Conferences:** 23 presentations in 12 countries: France - 1; U.S.A - 4; China -3; New Zealand - 1; Spain - 3; Canada-2; South Korea- 1; Switzerland- 2; Australia - 1; Mexico - 1; Japan -2; Itali - 1 and 3 keynote talks at international level.
- 8. Researcher mobilities and disciplines:** Total number of individuals engaged in the project is 34 from Mathematics, Computer Science, Psychology, Law, Sociology, Socio-Technical Studies Departments at Lancaster University and Cranfield University (including 12 PIs, Co-PI & Co-Is; 15 PDRAs, and 5 PhDs) from 9 countries (USA, UK, China, Sweden, Singapore, Turkey, Spain, Brazil, Mexico) with the next destination to 6 countries (USA, UK, Sweden, Brazil, France, Germany).
- 9. Demos:** ELSI Toolkit, Encoding Social Values in Autonomous Navigation: An Interactive Online Demonstration - see Case Study 1.
- 10. Other:** Embedding a TAS-S approach to RRI - Two workshops for all TAS-S researchers introducing the EPSRC RRI principles and considering their own contribution to RRI, and presenting their research in lay terms to promote public understanding.

3. Research overview

References

1. A. Pellicer, P. Angelov, N. Suri. Adversarial Image Threats in Autonomous Systems: A Survey; submitted for review ACM Computing Surveys, July 2023.
2. A. Perrusquia and W. Guo, 'Trajectory Inference of Unknown Linear Systems Based on Partial States Measurements,' in IEEE Transactions on Systems, Man, and Cybernetics: Systems.
3. A. Perrusquia, W. Guo, B. Fraser, Z. Wei, 'Uncovering Drone Intentions using Control Physics Informed Machine Learning,' Nature Communications Engineering, vol.3, Feb 2024.
4. A. Sogokon, B. Yuksek, G. Inalhan, N. Suri. Specifying Autonomous Systems Behaviour (arXiv preprint arXiv:2302.10087); submitted for review to Journal of Risk and Reliability.
5. A.E. af Wåhlberg, L. Dorn. (2024). The effects of Electronic Stability Control (ESC) on fatal crash rates in the United States, Journal of Safety Research, Volume 88, 217-229, <https://doi.org/10.1016/j.jsr.2023.11.008>.
6. af Wåhlberg A.E. & Dorn L (Submitted). Experience with automated systems; measurement issues.
7. af Wåhlberg A.E. & Dorn L. (Submitted). Meta-analysis of the safety effect of Electronic Stability Control.
8. af Wåhlberg, A. E., & Dorn, L. (submitted). Determinants of stated trust in automated systems; a comparative study of different predictions.
9. B. Yuksek, Z. Yu, N. Suri and G. Inalhan, 'Federated Meta Learning for Visual Navigation in GPS-denied Urban Airspace,' 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC), Barcelona, Spain, 2023.
10. C. Li and W. Guo, 'Soft Body Pose-Invariant Evasion Attacks against Deep Learning Human Detection,' 2023 IEEE Ninth International Conference on Big Data Computing Service and Applications, Athens, Greece, 2023
11. C. Li, S. Sun, Z. Wei, A. Tsourdos, W. Guo, 'Scarce Data Driven Deep Learning of Drones via Generalized Data Distribution Space,' Neural Computing and Applications, early access, Apr 2023
12. May-Chahal, C., Deville, J., Moffat, L., Guo, W., Tang, Y., Tsourdos, A. Encoding Social & Ethical Values in Autonomous Navigation: Philosophies Behind an Interactive Online Demonstration. ACM Proceedings: Second International Symposium on Trustworthy Autonomous Systems (TAS '24), Texas, USA.

3. Research overview

References

13. Michelin, A (2023). Distributed Epistemic Systems. STSMN (Science and Technology Studies in the Midlands and North) meeting hosted by the University of York (Sep 2023).
14. Moffat L., (2024) Ethics through the Wash: Narratives of Scandal in Autonomous Systems Research, Journal of Responsible Innovation (in press)
15. Moffat, L. (2023). Relational Approaches to Autonomous Systems Ethics. In Proceedings of the First International Symposium on Trustworthy Autonomous Systems (pp. 1-7).
16. O.V. Gunasekera, A. Sogokon, A. Gouglidis, N. Suri. Real Arithmetic in TLAPM, NASA Formal Methods NFM 2024 (in submission).
17. S. Qiu et al. (2024) 'Review of Physical Layer Security in Molecular Internet of Nano-Things,' in IEEE Transactions on NanoBioscience, vol. 23, no. 1, pp. 91-100, Jan.
18. W. Guo, Z. Wei, O. Gonzalez, A. Perrusquía and A. Tsourdos, 'Control Layer Security: A New Security Paradigm for Cooperative Autonomous Systems,' in IEEE Vehicular Technology Magazine, 2023
19. Y. Huang et al., 'Physical-Layer Counterattack Strategies for the Internet of Bio-Nano Things with Molecular Communication,' in IEEE Internet of Things Magazine, vol. 6, no. 2, pp. 82-87, June 2023
20. Z. Wei and W. Guo, 'Control Layer Security: Exploiting Unobservable Cooperative States of Autonomous Systems for Secret Key Generation,' in IEEE Transactions on Mobile Computing, 2024
21. Z. Wei, B. Li, W. Guo, 'Adversarial Reconfigurable Intelligent Surface Against Physical Layer Key Generation,' IEEE Transactions on Information Forensics & Security, vol.18, Apr 2023
22. Z. Wei, W. Hu, W. Guo, 'Explainable Adversarial Learning Framework on Physical Layer Secret Keys Combating Malicious Reconfigurable Intelligent Surface,' IEEE Transactions on Information Forensics & Security, under review, 2024.
23. Z. Yu, Y. Lu, N. Suri, MFSM: Mobility-aware Federated Split Meta Learning Framework in Distributed Environments, in progress.
24. Z. Yu, Y. Lu, N. Suri, RAFL: A Robust and Adaptive Federated Meta-Learning Framework Against Adversaries, in Proc. of IEEE MASS, 2023.
25. Z. Yu, Y. Lu, P. Angelov, N. Suri, PPFM: An Adaptive and Hierarchical Peer-to-Peer Federated Meta-Learning Framework, in Proc. of IEEE MSN, 2022. (Best Paper Award)

4. RS1 Activity Summary

RS1A: Development of specifications of threats and functional behaviour in Autonomous Systems

This research focused on the specification of Autonomous System (AS) threats along with the fundamental problems of formal verification and properties in models of AS embedded in a physical environment (such systems typically combine discrete and continuous behaviour and are thus examples of Cyber-Physical Systems).

Adversarial Image Threats: AS face growing adversarial threats in the image domain. These attacks exploit deep neural networks vulnerabilities, potentially causing significant malfunctions. This study reviews current adversarial attacks and defenses in AS, systematically analyzing the threat model within the AS stack, and outlining robust defense requirements. It discusses open issues and potential research directions and introduces the principles of a novel defense framework. Through an in-depth analysis of adversarial threats, it aims to steer future research towards more secure AS.

Publication: A. Pellicer, P. Angelov, N. Suri. Adversarial Image Threats in Autonomous Systems: A Survey; submitted for review ACM Computing Surveys, July 2023.

Specifications for Autonomous Systems: Specifying the intended behaviour of AS is becoming increasingly important but is associated with many challenges which have been the subject of numerous publications. A lack of specification makes it difficult to judge whether a system is functioning as intended, while imprecise or ambiguous specifications create difficulties from a regulatory standpoint (e.g. the system conforming an interpretation of a set of requirements published by a regulatory body). Formal requirements leave no scope for ambiguity and hold the promise of enabling truly trustworthy designs. We have compiled an overview of existing work on specifications of autonomous systems with a particular emphasis on formal specification, i.e. mathematically rigorous approaches to specification that require an appropriate formalism. The report features an overview of popular formalisms used in both academia and industry and provides useful examples and illustrations. Our report serves as a useful reference point providing a concise overview of the state of the art, along with current challenges and solutions in employing formal specifications in AS.

Publication: A. Sogokon, B. Yuksek, G. Inalhan, N. Suri. Specifying Autonomous Systems Behaviour (arXiv preprint arXiv:2302.10087); submitted for review to Journal of Risk and Reliability.

4. RS1 Activity Summary

Real Arithmetic in TLA+: Formally modelling and verifying properties (such as safety) of Cyber-Physical Systems requires a specification formalism and a logic. Lamport's Temporal Logic of Actions (TLA+) can serve this purpose and benefits from a mature set of tools developed over decades and successfully applied in large case studies. While TLA+ has mostly been used to specify and verify purely discrete systems, the formalism is expressive enough to allow specifications of continuous behaviour governed by ordinary differential equations (ODEs). In order to prove safety specifications about systems that feature ODEs, it is in practice necessary to find an appropriate inductive invariant which acts as an over-approximation of the set of reachable states of the system that contains no unsafe states described in the safety specification. Checking whether any unsafe states are present in the invariant requires real arithmetic. TLA+ supports real numbers, but up to now offered no support for automatically proving theorems of real arithmetic. The theory of (nonlinear polynomial) real arithmetic is supported by verification tools such as the Z3 Satisfiability Modulo Theory (SMT) solver, which already interfaces with the TLA+ Proof Manager (TLAPM) to automate proofs about integer arithmetic conjectures. We have extended the system to add support for real arithmetic conjectures, which can now be automatically discharged, thereby facilitating inductive invariant checking (and therefore safety verification) for CPS in TLA+.

Publication: O.V. Gunasekera, A. Sogokon, A. Gouglidis, N. Suri. Real Arithmetic in TLAPM, NASA Formal Methods NFM 2024 (in submission).

Synthesis of Barrier Certificates using Linear Programming Relaxations: Barrier certificates are functions that act as certificates of safety properties in models of computational and cyber-physical systems. Searching for barrier certificates in an inherently difficult problem, with traditional approaches relying on the use of sum-of-squares (SOS) relaxations and semidefinite programming (SDP), which are plagued by numerical issues. An alternative to SDP is to employ Linear Programming (LP) relaxations, which are numerically robust and provide certificates that can be formally verified by an external checker.

Publication: Current work in progress. Target conference submission: FM 2024, International Symposium.

4. RS1 Activity Summary

RS1B: Development of Frameworks for Federated ML in Autonomous Systems

This research has involved the progressive development of a framework for federated ML that can handle the AS-relevant aspects of data heterogeneity, adversarial data/model corruptions and incorporation of mobility aspects on the data.

Data heterogeneity: Distributed edge nodes within AS contain varied data distributions. To address the problem of data heterogeneity in AS, we proposed an adaptive and hierarchical Peer to Peer Federated Meta learning framework (PPFM). PPFM is a dynamic distributed machine learning approach, which uses multiple learning loops to dynamically self-adapt its own architecture to improve its training effectiveness for different generated data characteristics. In PPFM, a peer-to-peer FL framework is designed to enhance privacy of individual edge nodes and remove reliance on a fixed centralized server in a distributed environment.

Publication: Z. Yu, Y. Lu, P. Angelov, N. Suri, PPFM: An Adaptive and Hierarchical Peer-to-Peer Federated Meta-Learning Framework, in Proc. of IEEE MSN, 2022. (Best Paper Award)

Adversarial data/model corruptions: Due to distributed machine learning is vulnerable to adversarial attacks, we proposed a Robust and Adaptive Federated meta-Learning framework (RAFL) to adaptatively defend against a range of adversarial attacks. RAFL leverages a rule-based detection model and an online Variational AutoEncoder (VAE) detection model to identify and remove adversaries by comparing mean and standard deviation in a ranking matrix and reconstruction errors in VAE latent space. A similarity-based model aggregation method to further reduce the likelihood of uploading adversarial models from adversarial clients.

Publication: Z. Yu, Y. Lu, N. Suri, RAFL: A Robust and Adaptive Federated Meta-Learning Framework Against Adversaries, in Proc. of IEEE MASS, 2023.

4. RS1 Activity Summary

This work has been extended to the scenario of visual-inertial navigation. The most existing ML-based visual navigation solutions have weak adaptability to new environments and cannot be trained effectively on small-size datasets. We proposed a federated meta learning framework for visual navigation in GPS-denied urban airspace. A robust-by-design federated meta learning framework is proposed to achieve fast adaption to new environments and conditions. Adversarial attacks can be detected by learning multiple meta models and optimize a global meta model generalizability across vehicles that operate in different environments.

Publication: B. Yuksek, Z. Yu, N. Suri, G. Inalhan, Federated Meta Learning for Visual Navigation in GPS-denied Urban Airspace, in Proc. of AIAA/IEEE DASC, 2023.

Mobile AS nodes: The mobility of AS nodes presents difficulties, encompassing the risk of missing valid training data, dynamic fluctuations in the number of participating nodes, and degradation in the global model performance. This complexity is exacerbated by various characteristics of mobile nodes, including their node moving speed (slow or fast movement), relative direction (with or without route), and instances of overlap. To address the above challenges, we propose a Mobility-aware Federated Split Meta learning framework (MFSM) to divide moving edge nodes into different clusters for training personalized modes with the aid of meta learning. In MFSM, a federated split learning architecture is used to address the fast-changing data distribution, and a semantic communication-based clustering approach is utilized to quick assign edge nodes with non-IID dataset into different data distribution.

Publication: Z. Yu, Y. Lu, N. Suri, MFSM: Mobility-aware Federated Split Meta Learning Framework in Distributed Environments, in progress.

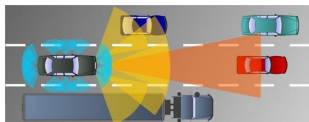
Real Arithmetic in TLA+

Towards proving properties in Cyber-Physical Systems

Lancaster University School of Computing and Communications

PhD Researcher: Ovin V.W. Gunasekera Supervisors: Dr. Antonios Gouglidis, Prof. Neeraj Suri

Towards Secure Usage of Autonomous Systems



Collision-Avoidance safety property of Autonomous Systems

- Autonomous systems are typically considered as Cyber-Physical Systems (CPSs)
- CPSs are considered safety-critical as undetected faults can result in catastrophic consequences of serious injury or loss of life, therefore establishing safety is crucial
- This research work enables automatically proving safety properties of CPSs via a verification tool - TLA+ Proof Manager

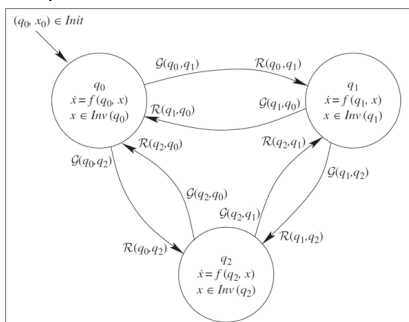
Formal Verification using TLA+

- TLA+ is a formal specification language developed by Leslie Lamport to model systems and programs
- TLA+ toolbox is a software tool which provides an IDE for writing and verifying TLA+ specifications
- This toolbox provides support for model checking via an explicit model checker (TLC) and deductive verification via the TLA+ Proof Systems (TLAPS)



Why TLA+

- TLA+ has gained attention from the academic community and industry and is used by major companies such as Amazon, Intel and Microsoft
- Based on Zermelo-Frankel set theory, the language enables specification and verification of wide range of systems from concurrent to distributed systems
- TLA+ is expressive enough to model hybrid systems i.e. systems which combine discrete and continuous behaviours (this includes CPSs)

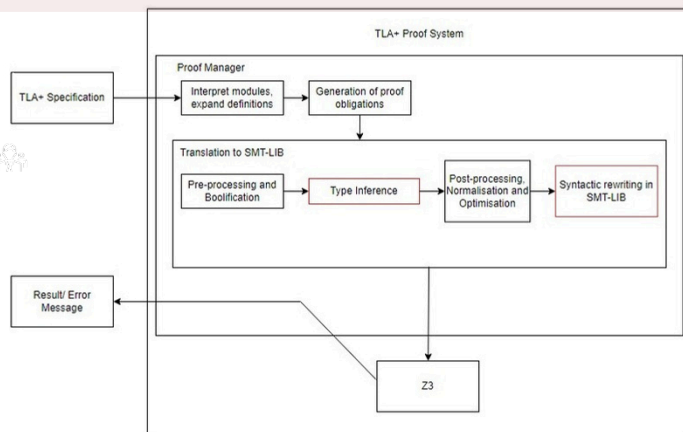


Example of a CPS modelled as a hybrid automaton

- TLAPS currently supports automatically proving theorems containing integer arithmetic
- Verification of properties in CPSs require modelling continuous state evolution and thus the representation of real numbers and real arithmetic is needed
- We extended the TLA+ Proof Manager to support proving real arithmetic conjectures to ultimately facilitate proving safety properties of CPSs



TLAPS Architecture



- TLAPS includes a proof manager which interfaces with several backend verifiers such as Z3, Isabelle and Zenon
- In extending the proof manager we enable automatically proving real arithmetic conjectures via Z3-SMT solver (SMT-LIB input) which facilitates proving of real arithmetic conjectures
- As highlighted in the TLAPS architecture diagram, we extended stages of the translation process from untyped TLA+ to multi-sorted SMT-LIB to interpret reals and real arithmetic

Enabling translation from TLA+ to SMT-LIB

Two types of translation

- Typed Encoding: Type inference algorithm and TLA+ type system assigns types to TLA+ expressions
- Untyped Encoding: Delegates type inference to SMT-solvers

If typed encoding fails to infer types to TLA+ expressions, type inference is delegated to SMT solvers

Extensions to typed encoding process

- Extended TLA+ type system by introducing type Real and typing rules to enable the interpretation of real arithmetic
- Constraint generation and constraint solving phases of the TLA+ type inference algorithm was extended to interpret real arithmetic
- Extensions to untyped encoding process
- Declared uninterpreted functions to embed SMT reals into a sort representing TLA+ values
- Introduced axioms to ensure soundness and consistency in translation during untyped encoding

Example of supported arithmetic operations in TLAPS and extended TLAPS

Operators	TLAPS		Extended TLAPS	
	Real	Int	Real	Int
Addition(+) Subtraction(-) Multiplication(*)	✓	✓	✓	✓
Integer division (\div) Modulus (%)	✓	✓	✗	✓
Division (/)	✓	✗	✓	✗
Range (..)	✓	✓	✓	✓
Unary minus (-)	✓	✓	✓	✓
Comparison (<, >, ≤, ≥)	✓	✓	✓	✓
Exponentiation (^)	✓	✗	✗	✗

We acknowledge Dr. Andrew Sogokon in this work.



This work is supported by the Engineering and Physical Sciences Research Council [grant number: EP/V026763/1]

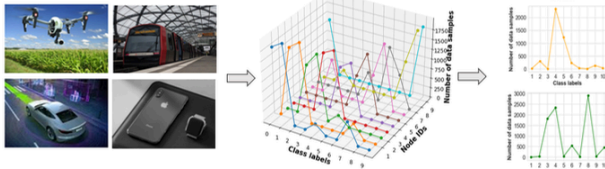
Machine Learning Frameworks for Autonomous System

Lancaster University, UK

Dr. Zhengxin Yu (z.yu8@lancaster.ac.uk) Prof. Neeraj Suri (neeraj.suri@lancaster.ac.uk)

Heterogenous Data in Autonomous System (AS)

- Distributed nodes in AS contain varied data distribution
- Centralized Machine Learning (ML) frameworks:
 - Reduce model accuracy
 - Privacy risk
 - Increase communication cost

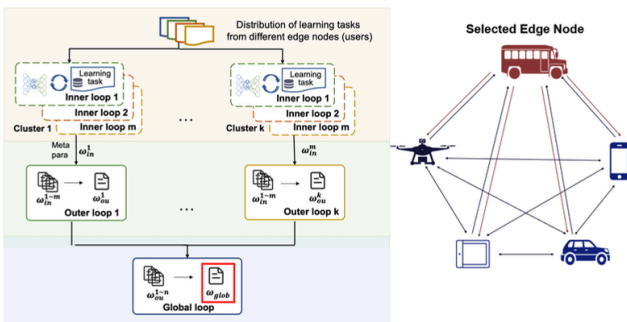


Adaptive and Hierarchical Peer-to-Peer Federated Meta-Learning Framework

Develop a hierarchal federated meta-learning framework to adaptively match the characteristics of heterogeneous data (PPFM)

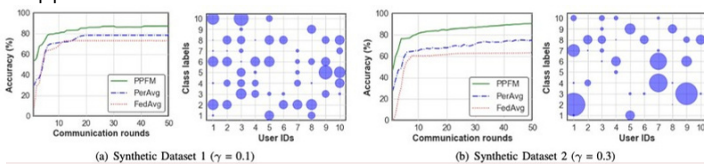
Contributions:

- A novel hierarchal meta-learning architecture
 - Generate multiple learning loops to match different data distribution
- A peer-to-peer federated learning approach
 - Ease reliance on the fixed central server
- A federated learning based data clustering method



Experimental results:

PPFM improves accuracy and efficiency over the state-of-art approaches

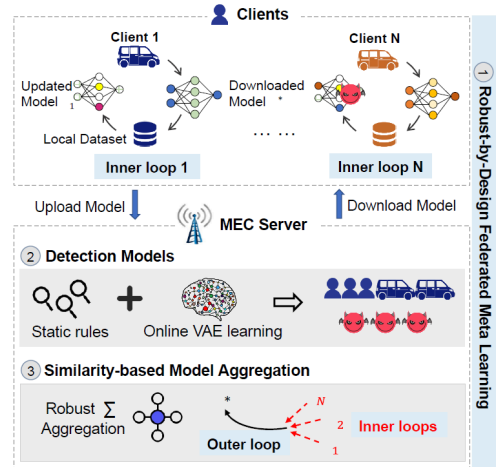


Heterogenous Data in Autonomous System (AS)

Develop a **robust and adaptive** federated meta-learning framework against adversaries (RAFL)

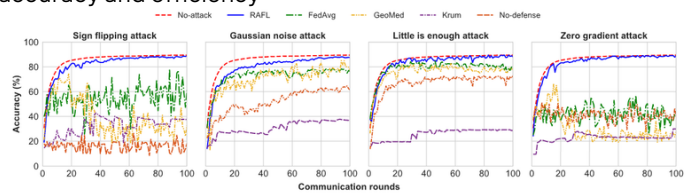
Contributions:

- A robust-by-design federated meta-learning architecture
 - Adaptively defend against a range of adversarial attacks.
- A composite rule-based and learning-based detection method
 - Identify adversaries via ranking domain and low-dimensional embeddings.
- An adaptive model aggregation method
 - Aggregate the global model by considering the degree of similarity between the meta-model and calculated mean model to resilience attacks.



Experimental results:

RAFL is robust by design and outperforms other baseline defensive methods against adversaries in terms of model accuracy and efficiency

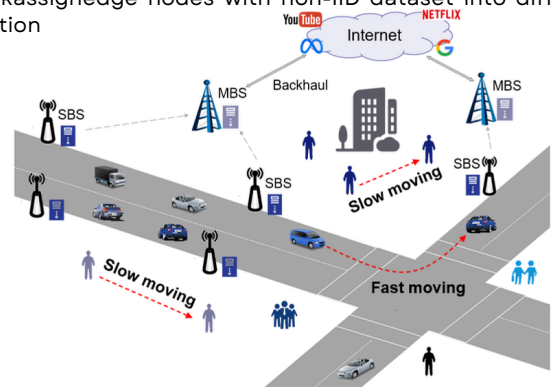


Mobility-aware Federated Learning Framework

Develop a **mobility-aware federated meta-learning framework** to reduce the impact of node mobility

Contributions:

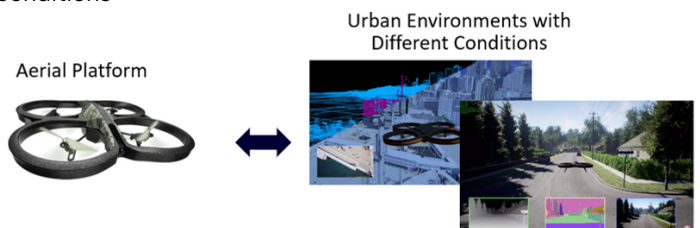
- A novel federated split learning architecture
 - Address the fast changing data distribution
- A semantic-based clustering approach
 - Quickassign edge nodes with non-IID dataset into different distribution



Case Study: Federated Meta Reinforcement Learning for UAV Navigation

Federated Learning-based Visual Odometry Framework

- Combining the AI-based solutions with classical filter-based approach
- Utilising RAFL framework to improve pose estimation accuracy
- Aggregating models trained in different environments and conditions



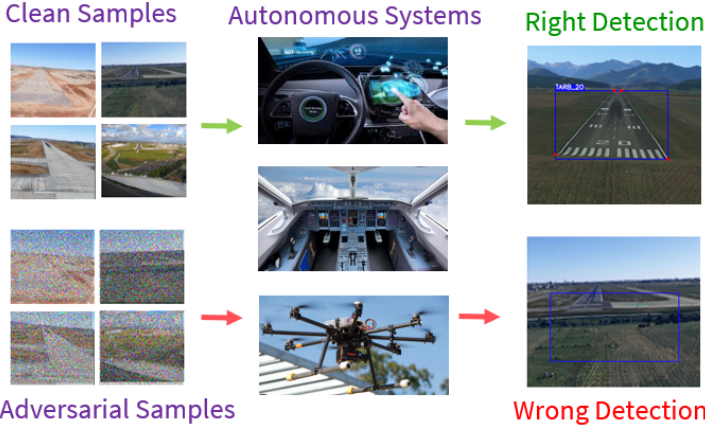
Rethinking Self-supervised Learning for Cross-domain Adversarial Image Recovery

Lancaster University

Research Fellow: Dr. Yi Li Investigators: Prof. Plamen Angelov, Prof. Neeraj Suri

Adversarial Attacks to Autonomous Systems

Autonomous Systems (AS) are usually embodied as Cyber-Physical Systems (CPS) in which adversarial attacks can lead to catastrophic consequences, such as loss of life or serious injury, thus many autonomous systems are **safety-critical**.



Self-supervised Learning

What is self-supervised learning (SSL):

- Unlabeled data is processed to obtain useful representations that can help with downstream learning tasks.
- An intermediate form of unsupervised and supervised learning.

Why we need SSL-based adversarial attack recovery?

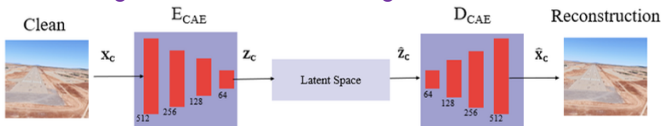
- Supervised training of the networks requires large sets of labelled paired data. However, these data is difficult or expensive to obtain.
- A trained model may suffer from performance degradation when deployed in previously unseen conditions e.g., a mismatch of attacks and datasets between the training and testing datasets.

What do we propose in this work?

- We propose the clean image autoencoder (CAE) to learn the latent representations of clean images.
- We propose the adversarial image autoencoder (AAE) to learn a shared latent space between the unpaired clean images and adversarial images to boost the generalization ability.
- The input of two autoencoders are clean images and adversarial images, respectively. However, they are unpaired, i.e., they are randomly selected different domains (datasets and attack algorithms).

Proposed Framework

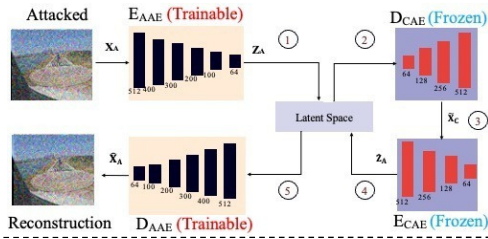
Clean image autoencoder training



- The clean images X_c from the public landing runway dataset are fed into the CAE to learn the features Z_c in the latent space.
- In CAE, both ECAE and DCAE consist of four 1-D convolutional layers. In ECAE, the size of the hidden dimension decreases sequentially from 512 \rightarrow 256 \rightarrow 128 \rightarrow 64. Accordingly, the dimension of the latent space is set to 64, with the stride of 1 and the kernel size of 7 used for the convolutions. Different from ECAE, the decoder DCAE scale up the latent dimensions sequentially.

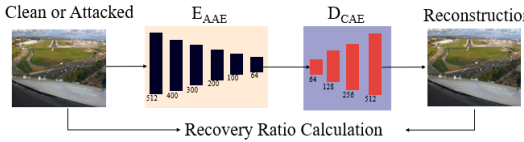
Proposed Framework

Adversarial image autoencoder training



- The weights of the CAE are frozen in this stage.
- The AAE learns a shared latent space between clean images and adversarial images.

Test stage



- The trained EAEE and DCAE are combined as the final model

Experimental Settings

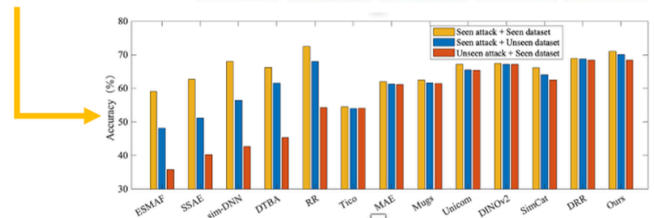
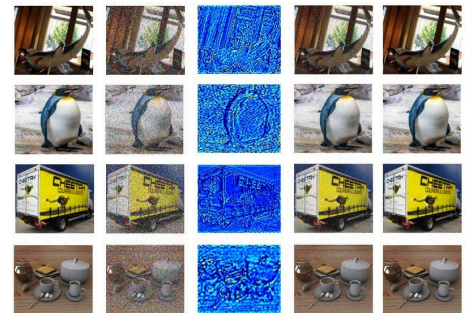
- CAE Training: 10k images from the COCO dataset
- AAE Training: 40k images from the CIFAR-10 dataset
- Test: 10k images from the ImageNet-R dataset
- Backbones: CNN
- Attack algorithms: FGSM, PGD, SSAH, DeepFool, BIM, CW, JSMA

Experimental Results

	Recovery Ratio (%)								
	Clean	FGSM	PGD	SSAH	DeepFool	BIM	CW	JSMA	Avr
ESMAF	70.8	52.7	67.5	62.9	39.7	35.9	37.2	41.0	51.0
SSAE	74.0	58.5	67.0	69.2	41.5	39.4	41.6	41.3	54.1
Sim-DNN	76.2	60.7	72.3	71.0	44.8	46.7	49.9	50.2	59.0
DTBA	79.2	59.2	75.5	74.9	51.4	53.8	56.0	59.9	63.8
RR	86.5	62.7	79.0	76.2	67.1	58.7	60.9	71.3	70.3
TiCo	74.5	53.6	68.6	65.2	45.1	44.5	43.1	57.9	56.6
MAE	82.2	59.6	75.5	74.4	54.2	50.3	51.4	62.8	63.8
Mugs	83.4	57.2	75.9	76.7	56.0	51.1	50.8	64.3	64.4
Unicom	86.4	59.8	76.2	79.3	61.0	55.5	58.4	68.2	68.1
DINOv2	87.5	61.6	79.4	78.3	64.5	57.1	57.9	71.6	69.7
SimCat	85.1	58.0	75.2	77.0	56.4	56.5	55.3	69.6	66.6
DRR	87.2	64.8	79.6	78.2	66.9	60.7	60.1	70.3	71.0
Ours	87.9	65.9	80.0	79.7	69.1	61.5	61.8	72.4	72.3

Visualizations

- Results on the Image-R dataset.
- Supervised: ESMAF, SSAE, sim-DNN, DTBA, RR
- Self-supervised: Tico, MAE, Mugs, Unicom, DINOv2, SimCat, DRR



Ongoing and Future Works

- The proposed framework is potentially applied in other downstream tasks, e.g., road condition detection.
- Ablation study of the proposed algorithm will be provided.

UNICAD: A Unified Approach for Attack Detection, Noise Reduction and Novel Class Identification

Lancaster University

Researcher: Alvaro Lopez Pellicer Supervisors: Prof. Plamen Angelov, Prof. Neeraj Suri

Challenges for Autonomous Systems

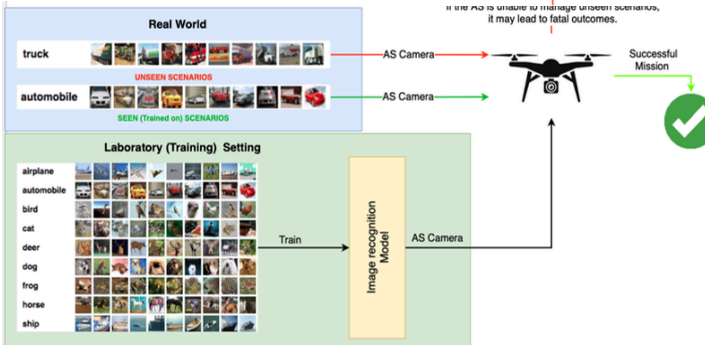
Autonomous systems face numerous challenges in their operation due to the uncertain and dynamic multi-layer attack surfaces

Critical Challenges

Accurate operation: Autonomous Systems (AS) consist of complex ensembles of interconnected components including sensors, actuators, communication modules and control algorithms, that collaboratively perform tasks with minimal to no human intervention. Often, these systems rely on image sensing for perception and decision-making in the physical environment. Each discrete component needs to individually perform at the requisite level of accuracy in order to result in a collectively stable AS operation.

Safety: Autonomous Systems are safety-critical and increasingly utilize Deep Neural Networks for multiple tasks (DNNs). Adversarial attacks are one of the most critical challenges for DNNs and AS. These attacks can take various forms, such as data poisoning, model inversion, or evasion, and can have serious consequences for the safety, reliability, and privacy.

Unknown scenarios: DNNs are often trained in a set of known scenarios. For instance, they may be trained to identify different objects in aerial images, however, when a new object appears in which the systems hasn't been trained on, often it would be misclassified. AS need to be able to detect and handle unknown scenarios, failure to do so could lead to catastrophic consequences.



Why do we need a unified solution?

Autonomous Systems need to:

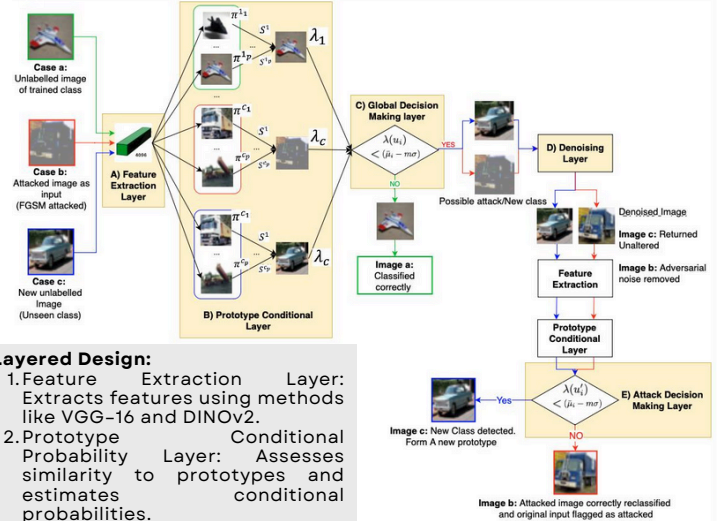
- Operate correctly under trained scenarios
- Be able to recognize (detect) Adversarial Attacks in real time
- React to adversarial attacks often by mitigating their impact
- Detect unseen scenarios out of the scope of original training
- React to unseen scenarios

Our proposed solution (UNICAD) compared to others:

Method	Unseen Class detection	Attack detection	Attack Recovery
Previous work on unseen class detection (xClass)	✓	✗	✗
Previous work on attack detection (simDNN)	✗	✓	✗
Denoising Autoencoders (DAE)	✗	✗	✓
Ours	✓	✓	✓

UNICAD Framework Overview

A novel architecture integrating state-of-the-art techniques for efficient adversarial attack detection, noise reduction, and novel class recognition



Layered Design:

- Feature Extraction Layer:** Extracts features using methods like VGG-16 and DINOv2.
- Prototype Conditional Probability Layer:** Assesses similarity to prototypes and estimates conditional probabilities.
- Global Decision-Making Layer:** Classifies input as an existing class, new class, or adversarial attack.
- Denoising Layer:** A denoising autoencoder that removes adversarial noise while preserving the integrity of clean inputs.
- Attack Decision Making Layer:** Re-evaluates denoised images to determine if they represent a new class or an attack.

Key Features

- Effective in detecting adversarial attacks and data concept drifts (Unseen scenarios).
- Reduces Adversarial noise using advanced Denoising Autoencoders.
- Identifies new classes using similarity-based neural networks.
- Maintains performance in seen and not attacked scenarios (Normal scenarios)

Results and Discussion

Scenario	Accuracy (%)					
	UNICAD (VGG-16 FE)	UNICAD (DINOv2 FE)	xClass (VGG-16 FE)	Traditional DAE (Defence)	VGG-16 (No defence)	DINOv2 (No defence)
Clean	80.86	92.93	80.86	81.4	92.0	97.63
PGD ($\epsilon = 0.01$)	74.81	77.77	49.3	61.10	0.05	56.8
PGD ($\epsilon = 0.3$)	72.63	82.29	14.2	61.10	32.12	0.7
FGSM ($\epsilon = 0.01$)	70.01	77.37	49.0	63.00	31.06	58.9
FGM ($\epsilon = 0.03$)	64.9	76.02	47.1	61.00	22.56	17.3
FGM ($\epsilon = 0.3$)	73.09	81.10	15.6	57.10	0.11	12.4
C&W (L2 norm)	73.2	79.33	0.6	36.70	0.00	0.8
Unseen Class detection	62.30	83.38	62.30	0.00	0.00	0.00

Experimental Setup

- Framework Validation:** CIFAR-10 datasets to evaluate UNICAD's robustness.
- Unseen class setting:** UNICAD and comparative methods trained on CIFAR-10 classes 0-8, leaving class 9 (trucks) unseen.
- Performance Assessment Criteria:** Analysing the classification accuracy of comparative methods in clean settings, against FGSM, PGD, C&W attacks and unseen scenarios. Accuracy measured on the CIFAR-10 testing dataset and CIFAR-9 for unseen class detection.
- Unseen Class Detection metric:** $\text{Detection}(\%) = \frac{TP + TN}{TP + FP + TN + FN} \times 100$

Key Results

- UNICAD with VGG-16 FE: Clean image classification slightly higher than DSA, over 70% accuracy in adversarial attacks (FGSM, PGD, C&W), comparable unseen class detection to xClass.
- UNICAD with DINOv2 FE: Enhanced feature extraction leading to superior performance, significant lower accuracy drop in adversarial attacks, robust in unseen class detection.
- Performance Comparison: Demonstrates robustness in adversarial attack scenarios, effective in unseen class detection, balanced approach to classification accuracy and security against attacks.

Future work

- Exploring Latent Space Similarities: Investigate the similarity of different classes in the latent space and its relevance, especially for closely related classes like trucks and automobiles.
- Optimizing Denoising Layer: Work on optimizing the denoising layer to better adapt to a variety of changing and evolving adversarial attacks.
- Exploring Real-world scenarios: Test UNICAD on closer to real-world scenarios such as simulations or more realistic datasets such as LARD dataset.

Autonomous Systems: Specification and Verification

Lancaster University School of Computing and Communications

Andrew Sogokon (a.sogokon@lancaster.ac.uk)
Prof Neeraj Suri (PI)



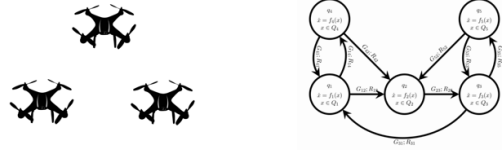
Engineering and Physical Sciences Research Council



UKRI Trustworthy Autonomous Systems Hub

Cyber-Physical Autonomous Systems

- Systems that interact with a physical environment are *cyber-physical systems (CPS)*.
- Continuous dynamics in CPS is usually described using **differential equations**.



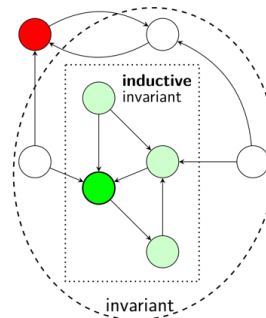
- Formal models of CPS involve **real numbers** and formal verification **requires real arithmetic**.



Formal Modelling and Verification in TLA+

Temporal Logic of Actions

- Lampert's Temporal Logic of Actions was designed to enable formal modelling and verification of concurrent systems. It enjoys excellent tool support in the form of the **TLA+ Toolbox** and has been successfully applied in industry.
- Formally proving safety specifications of discrete transition systems is typically done by finding an appropriate **invariant**.



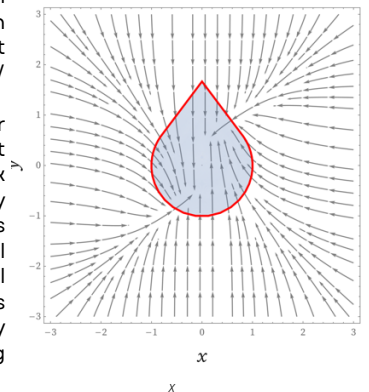
Inductive Invariants

An invariant is a set of states that:

- It includes all the initial states (as described in the safety specification).
- It does not include any of the unsafe states.
- The unsafe states are not reachable from the initial states.

An invariant is inductive if there are no transitions out of the invariant.

- A corresponding notion to an inductive invariant in continuous systems is that of a positively invariant set / continuous invariant.
- Recent work in computer science has established that it is decidable to check whether a set is positively invariant (provided it is described using polynomial functions). This requires real arithmetic. This result makes it possible to perform safety verification without having to solve the ODEs.

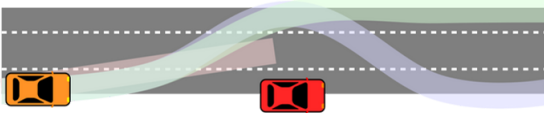


Automating Real Arithmetic in TLA+

- TLA+** supports real numbers (which are required for modelling and verifying CPS, especially in checking continuous invariants).
- However, the proof system currently lacks support for automatic proofs of first-order real arithmetic sentences (e.g. $\forall x, y \in \mathbb{R}. 2x^2 + (xy - y)^2 \geq -1$).
- The **TLA+ Proof Manager (TLAPM)** has now been extended to support nonlinear real arithmetic (O. V. Gunasekera et al.) – a step towards safety verification of CPS using TLA+.

Specifications for Autonomous Systems

- Specifications are descriptions of what a system should (or should not) do.
- A large source of specifications for AS comes from **regulations** (e.g. the *Highway Code* for terrestrial vehicles, or the *Rules of the Air* for aerial vehicles).
- Regulations written in natural language (e.g. English prose) can be imprecise and subject to various interpretations.
- E.g. "When changing the lane to the left lane during overtaking, no following road user shall be **endangered**" (Rizaldi et al., 2017).



Formal Specifications

- A **formal model** of the system provides a precise description of the dynamics.
- A **formal specification** can be verified against a formal model.
- Mission specifications can often be stated in formal logic (such as various **temporal logics**) and can incorporate safety and liveness requirements:

$$-\square_{[0,\infty)} \text{dist}(\text{ownship}, \text{intruder}) \geq d_{\min} \quad (\text{collision avoidance})$$



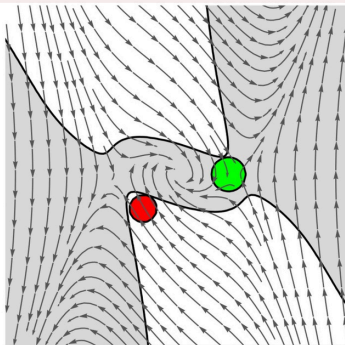
$$-\diamond_{[0,T]} \text{Target} \wedge (\square_{[0,T]} \text{Safe}) \quad (\text{reach-avoid})$$



Safety Specification and Verification

Safety Specifications

- A safety specification for a given system requires two elements:
 - 1 - A description of the possible initial states from which the system may begin its operation.
 - 2 - A description of undesirable (i.e. unsafe) states into which the system must never transition.
- Safety verification** is concerned with proving a safety specification, i.e. rigorously demonstrating that a system may never transition into any of the unsafe states provided that it starts operating from one of the specified initial states.



5. RS2 Activity Summary

RS2 A-B: Development of GNSS-Denied Navigation

Urban air mobility (UAM) is one of the most critical research areas which combines vehicle technology, infrastructure, communication, and air traffic management topics within its identical and novel requirement set. Navigation system requirements have become much more important to perform safe operations in urban environments in which these systems are vulnerable to cyber-attacks. Although the global navigation satellite system (GNSS) is a state-of-the-art solution to obtain position, navigation, and timing (PNT) information, it is necessary to design a redundant and GNSS-independent navigation system to support the localization process in GNSS-denied conditions. Recently, Artificial Intelligence (AI)-based visual navigation solutions are widely used because of their robustness against challenging conditions such as low-texture and low-illumination situations. However, they have weak adaptability to new environments if the size of the dataset is not sufficient to train and validate the system.

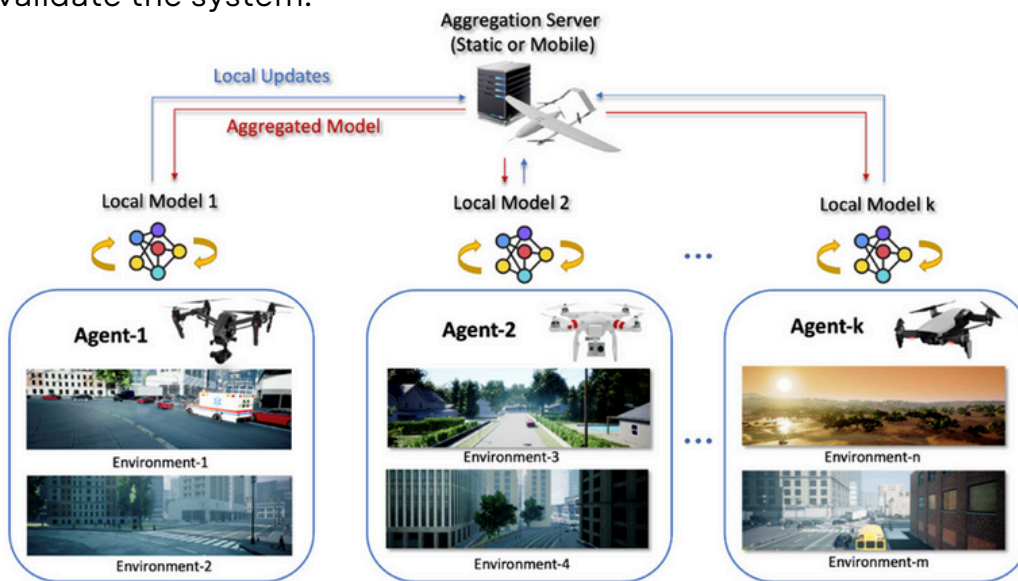


Figure 1: General overview of Federated Learning for Visual Odometry

5. RS2 Activity Summary

To address these problems, federated meta learning can help fast adaptation to new operation conditions with small dataset, but different visual sensor characteristics and adversarial attacks add considerable complexity in utilizing federated meta learning for navigation. Therefore, we proposed a robust-by-design Federated Meta Learning based visual odometry (see Figure 2.1) algorithm to improve pose estimation accuracy, dynamically adapt to various environments by using differentiable meta models and tuning its architecture to defense against cyber-attacks on the image data. In this proposed method, multiple learning loops (inner-loop and outer-loop) are dynamically generated. Each vehicle utilizes its collected visual data in different flight conditions to train its own neural network locally for a particular condition in the inner loops. Then, vehicles collaboratively train a global model in the outer loop which has generalizability across heterogeneous vehicles to enable lifelong learning.

Publications: B. Yuksek, Z. Yu, N. Suri and G. Inalhan, 'Federated Meta Learning for Visual Navigation in GPS-denied Urban Airspace,' 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC), Barcelona, Spain, 2023

Today's autonomous systems heavily rely on the current Global Navigation Satellite System (GNSS) solution for accurate position and speed tracking. These systems can be easily spoofed or jammed by attacker making them unreliable for autonomous systems (ASs). These attacks are harmful and difficult to detect. The GNSS system should be supported by utilizing multi-sensor pose estimation algorithms not only to detect the attacks but also to provide safety for the vehicle. To increase safety of the vehicles, we designed AI-aided Visual-Inertial navigation system to support the GNSS in the presence of spoofing attacks.

We utilized KITTI dataset for ground vehicle to develop and test our algorithm. Images from stereo camera are used to create depth measurement using semi global block matching (SGBM) algorithm. Depth measurement from stereo cameras is complemented with 360-degree LIDAR sensor to fill missing pixel and to provide multi sensor modality thus increasing the system overall safety. Front view of vehicle in grey scale, depth measurement of each pixel from stereo camera and inertial measurement unit (IMU) are used in visual inertial navigation (VIN) algorithm to estimate vehicle pose.

5. RS2 Activity Summary

The GPS spoofing system is designed to detect when an attacker jams or spoofs the GPS signals. VIN pose estimation is compared with GPS measurements to check for an attack. Once an intrusion is detected, this is relayed to the data fusion to switch from GPS to VIN position/velocity estimation.

We have developed AI aided visual inertial navigation algorithms providing robust navigation solution accuracy in austere environments subject to both the loss of GNSS and GNSS spoofing. We tested our visual inertial navigation (VIN) for GPS-denied environments and GPS spoofing detection algorithm on KITTI dataset to verify that the pose estimation can be made in data fusion step without affecting the vehicle guidance system. The results we obtained demonstrate robust spoofing detection and autonomous navigation in challenging environments, providing a cornerstone capability towards trustworthy autonomous systems.

In the near future, we anticipate seeing thousands of drones taking to the skies, navigating through urban and rural landscapes, including more challenging environments. However, the widespread use of drones brings with it a set of problems. Drones may run into emergencies such as low battery levels, issues with trust, and even potential manipulation of their GPS signals. Even those equipped with visual inertial navigation capabilities must confront the critical task of autonomously identifying and navigating towards safe landing locations. This intricate interplay of technology and environmental demands underscores the need for robust systems that can adeptly respond to emergent issues, ensuring the safe and effective operation of these unmanned aerial vehicles.

Addressing these challenges is crucial, and to this end we have developed an autonomous safe landing system to ensure the smooth and secure operation of these unmanned aerial vehicles in diverse scenarios. This system works in parallel to support VIN system enhancing overall trustworthiness. The system uses similar sensors to VIN such as stereo and LIDAR sensors, thus increasing mission safety without increasing payload weight and volume.

5. RS2 Activity Summary

Within our design, we used a 360-degree LIDAR and a stereo camera to obtain point cloud of the overflown ground. To reduce the error due to constant altitude change in air vehicles, the point cloud measurements are corrected by using the vehicle current orientation from inertial sensor. The availability of safe landing location is evaluated based on slope and roughness of the terrain. Each sensor measurement is divided into grids and the ones that pass the safety checks and closest to drone current location is selected for landing. In normal mission, the drone continuously scans the ground for suitable landing locations in case an emergency arises. When the emergency safe landing is requested either by flight computer or human operator, the algorithm autonomously diverts its path to available safe landing sites. To demonstrate the concept, we built a custom drone platform with stereo and LIDAR sensor suites and necessary computation power within SWAP-C constraints for flight testing. The developed safe landing location identification algorithm verified and validated in outdoor flight tests. As a result, we have developed and demonstrated an AI-aided landing location identification system to enhance system trustworthiness, particularly supporting Visual Inertial Navigation (VIN) and improving overall safety in diverse environments. The outdoor demonstration flight tests are conducted to validate the algorithm. This capability is integral in urban and rural airspace operations in which reliability and safety are provided while demonstrating a trustworthy autonomous system.

5. RS2 Activity Summary

RS2 B-C: Development of Secure Control & Communications

This research was joint between the control and information layer as a novel form of physics-driven digital encryption in swarm AS. Swarm AS often conduct cooperative control to ensure safe navigation. Embedded within the centralized or distributed coordination mechanisms are a set of observations, unobservable states, and control variables.

Security of data transfer between AS is crucial for safety, and both cryptography and physical layer security (PLS) methods have been used to secure communication surfaces—each with its drawbacks and dependencies. Existing computational complexity-based cryptography does not have information-theoretical bounds and poses threats to superior computational attackers. This post-quantum cryptography issue indeed motivated the rapid advances in using common physical layer properties to generate symmetrical cipher keys (known as PLS). However, PLS remains sensitive to attackers (e.g., jamming) that destroy its prerequisite wireless channel reciprocity.

When AS are in cooperative tasks (e.g., rescuing searching, and formation flight), they will behave cooperatively in the control layer. Inspired by this, we propose a new security mechanism called control layer security (CLS – see Figure 2), which exploits the correlated but unobservable states of cooperative ASs to generate symmetrical cipher keys.

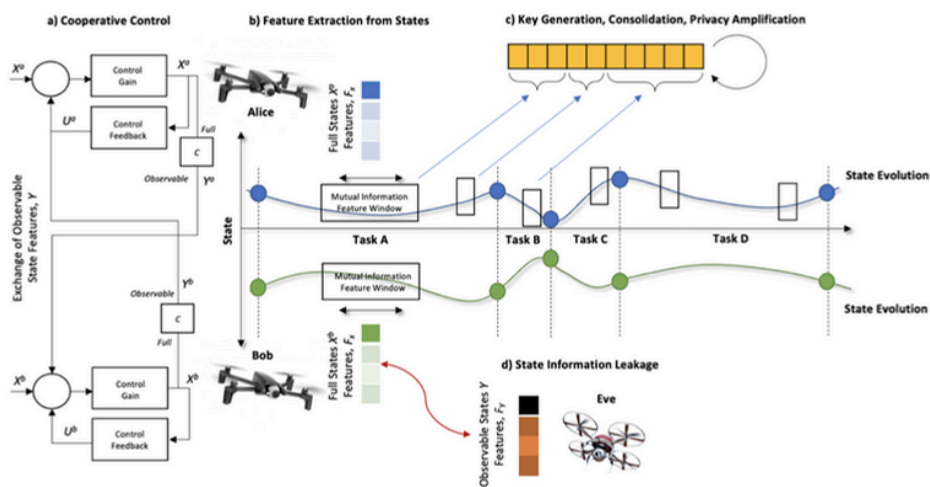


Figure 2: Control Layer Security.

5. RS2 Activity Summary

This idea is then realized in the linearized UAV cooperative control scenario. The theoretical correlation coefficients between Alice's and Bob's states are computed, based on which common feature selection and key quantization steps are designed. The results from simulation and real UAV experiments show:

- i) an approximately 90% key agreement rate is achieved, and
- ii) even an Eve with the known observable states and systems fails to estimate the unobservable states and the secret keys relied upon, due to the multiple-to-one mapping from unobservable states (pitch, roll and yaw angles) to the observable states (3D trajectory).

This demonstrates CLS as a promising candidate to secure the communications of AS, especially in the adversarial radio environment with attackers that destroys the prerequisite for current PLS.

Publications:

Z. Wei and W. Guo, 'Control Layer Security: Exploiting Unobservable Cooperative States of Autonomous Systems for Secret Key Generation,' in IEEE Transactions on Mobile Computing, 2024

W. Guo, Z. Wei, O. Gonzalez, A. Perrusquía and A. Tsourdos, 'Control Layer Security: A New Security Paradigm for Cooperative Autonomous Systems,' in IEEE Vehicular Technology Magazine, 2023

5. RS2 Activity Summary

RS2 A-B-C: Development of Inferring the Security Risk of an Uncooperative AS

This research was joint between the mission, control and information layer as a novel form of physics-informed inverse learning against uncooperative or suspicious AS.

Unmanned Autonomous Vehicle (UAV) or drones are increasingly used across diverse application areas. Uncooperative drones do not announce their identity/flight plans and can pose a potential risk to critical infrastructures. Understanding a drone's intention is important to assigning risk and executing countermeasures. Intentions are often intangible and unobservable (see Figure 2.3), and a variety of tangible intention classes are often inferred as a proxy. However, inference of drone intention classes using observational data alone is inherently unreliable due to observational and learning bias.

Here, we developed a control-physics informed machine learning (CPhy-ML) that can robustly infer across intention classes. The CPhy-ML couples the representation power of deep learning with the conservation laws of aerospace models to reduce bias and instability. The CPhy-ML achieves a 48.28% performance improvement over traditional trajectory prediction methods. The reward inference results outperform conventional inverse reinforcement learning approaches, decreasing the root mean squared spectral norm error from 3.3747 to 0.3229.

Publications: A. Perrusquia, W. Guo, B. Fraser, Z. Wei, 'Uncovering Drone Intentions using Control Physics Informed Machine Learning,' Nature Communications Engineering, vol.3, Feb 2024

5. RS2 Activity Summary

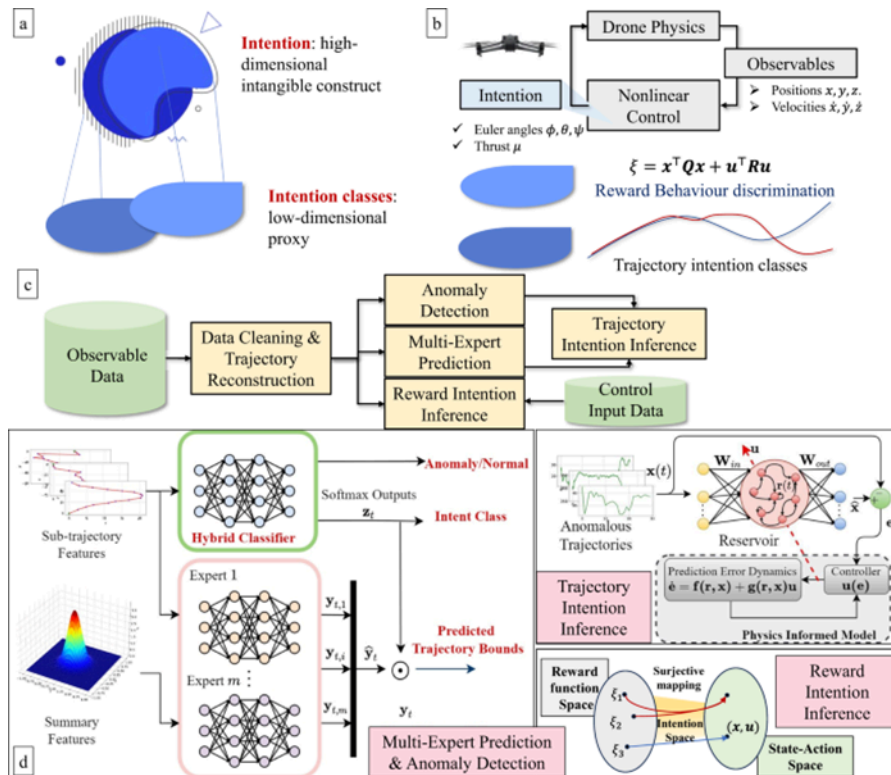


Figure 3: Intention Prediction for Autonomous Systems.

An accurate trajectory inference algorithm is required for monitoring and early detection of autonomous misbehaviour and to take relevant countermeasures. This article presents a trajectory inference algorithm based on a CLOE approach using partial states measurements. The approach is based on a physics informed state parameterization that combines the main advantages of state estimation and identification algorithms. Noise attenuation and parameter estimates convergence are obtained if the output trajectories fulfil a persistent excitation condition. Known and unknown desired reference/destination cases are considered. The stability and convergence of the proposed approach are assessed via Lyapunov stability theory under the fulfilment of a persistent excitation condition. Simulation studies are carried out to verify the effectiveness of the proposed approach.

Publications: A. Perrusquía and W. Guo, 'Trajectory Inference of Unknown Linear Systems Based on Partial States Measurements,' in IEEE Transactions on Systems, Man, and Cybernetics: Systems

5. RS2 Activity Summary

RS2 C: Dealing with Sparse, Biased, and Adversarial Training Data for Uncooperative Target Recognition

Increased drone proliferation in civilian and professional settings has created new threat vectors for airports and national infrastructures. The economic damage for a single major airport from drone incursions is estimated to be millions per day. Due to the lack of balanced representation in drone data, training accurate deep learning drone detection algorithms under scarce data is an open challenge. Existing methods largely rely on collecting diverse and comprehensive experimental drone footage data, artificially induced data augmentation, transfer, and meta-learning, as well as physics-informed learning. However, these methods cannot guarantee capturing diverse drone designs and fully understanding the deep feature space of drones.

Here, we show how understanding the general distribution of the drone data via a generative adversarial network (GAN) and explaining the under-learned data features using topological data analysis (TDA) can allow us to acquire under-represented data to achieve rapid and more accurate learning. We demonstrate our results on a drone image dataset, which contains both real drone images as well as simulated images from computer-aided design. When compared to random, tag-informed, and expert-informed data collections (discriminator accuracy of 94.67%, 94.53% and 91.07%, respectively, after 200 epochs), our proposed GAN-TDA-informed data collection method offers a significant 4% improvement (99.42% after 200 epochs). We believe that this approach of exploiting general data distribution knowledge from neural networks can be applied to a wide range of scarce data open challenges.

Publications: C. Li, S. Sun, Z. Wei, A. Tsourdos, W. Guo, 'Scarce Data Driven Deep Learning of Drones via Generalized Data Distribution Space,' Neural Computing and Applications, early access, Apr 2023

5. RS2 Activity Summary

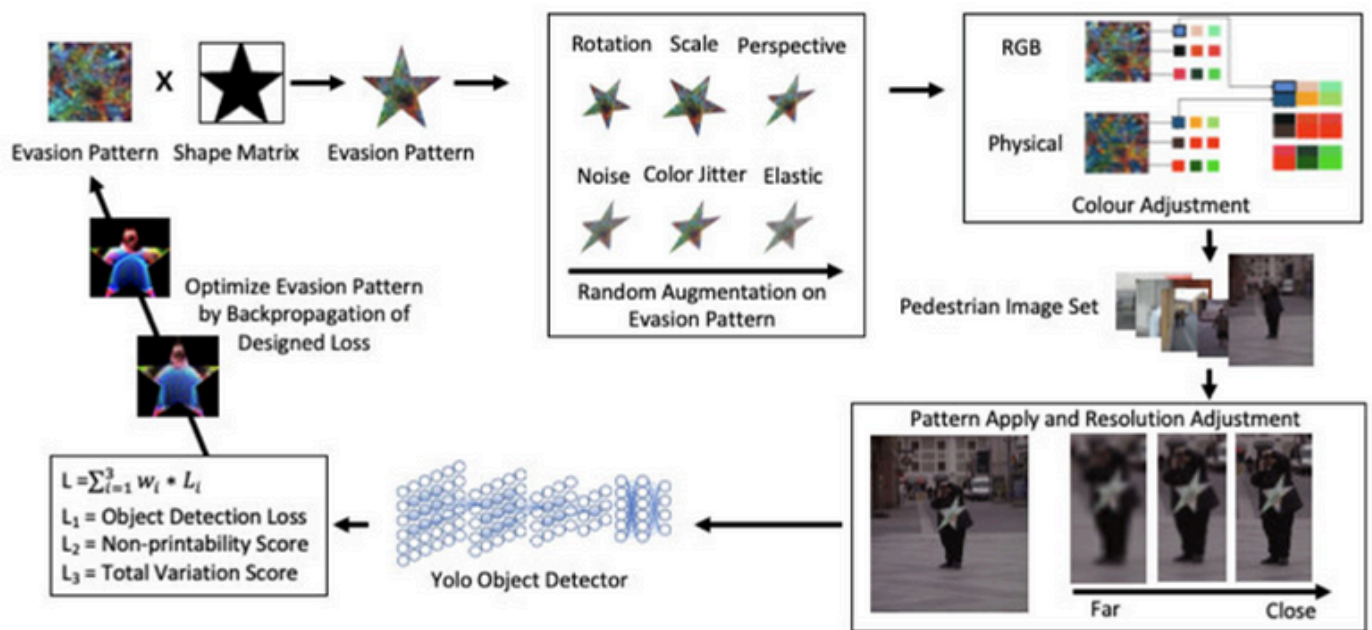


Figure 4: The process of training evasion patterns with 3D deformation

Evasion attacks on deep neural networks (DNN) use manipulated data to let targets evade detection and/or classification across a wide range of DNNs. Most existing evasion attacks focus on planar images (e.g., photo, satellite imaging) and ignore the distortion of evasion in practical attacks (e.g., object rotation, deformation). Here, we build evasion patterns for soft-body human stakeholders, where patterns are designed to consider body rotation, fabric stretch, printability, and lighting variations. We show that these are effective and robust to different human poses. This still represents a significant threat to safety of autonomous vehicles and adversarial training should consider this new area.

Publications: C. Li and W. Guo, 'Soft Body Pose-Invariant Evasion Attacks against Deep Learning Human Detection,' 2023 IEEE Ninth International Conference on Big Data Computing Service and Applications, Athens, Greece, 2023

5. RS2 Activity Summary

RS2 C: Identifying and Countering the Risks of Adversarial Meta-Surfaces for Secure Communications

The development of reconfigurable intelligent surfaces (RIS) has recently advanced the research of physical layer security (PLS). Beneficial impacts of RIS include but are not limited to offering a new degree-of-freedom (DoF) for key-less PLS optimization and increasing channel randomness for physical layer secret key generation (PL-SKG). However, there is a lack of research studying how adversarial RIS can be used to attack and obtain legitimate secret keys generated by PL-SKG.

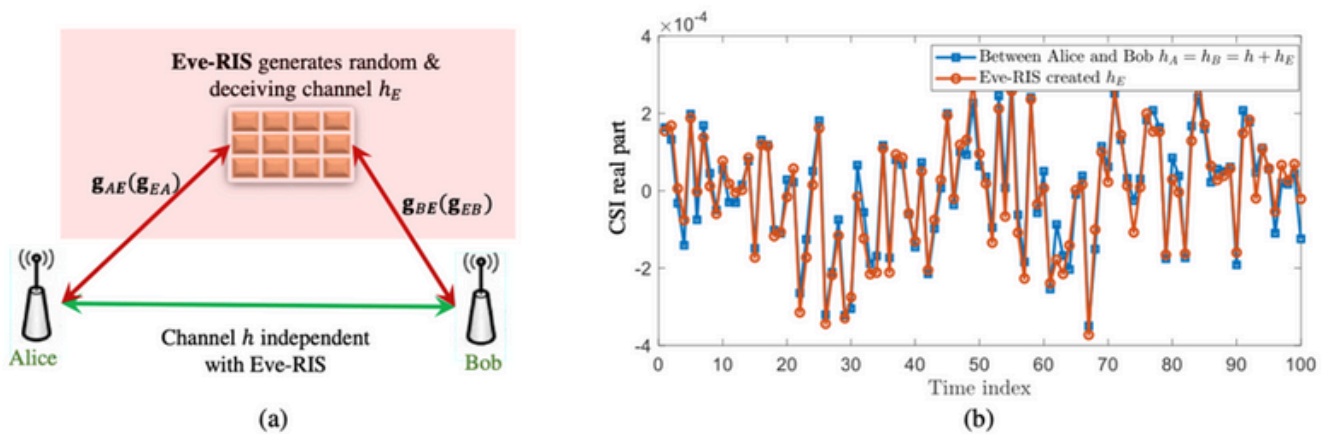


Figure 5: Sketch of Eve driven RIS.

In this red-teaming work, we show how an Eve-controlled adversarial RIS (Eve-RIS), by inserting into the legitimate channel a random and reciprocal channel, can partially reconstruct the secret keys from the legitimate PL-SKG process. To operationalize this concept, we design Eve-RIS schemes (Figure 2.5) against two PL-SKG techniques used: (i) the CSI-based PL-SKG, and (ii) the two-way cross multiplication-based PL-SKG. The channel probing at Eve-RIS is realized by compressed sensing designs with a small number of radiofrequency (RF) chains. Then, the optimal RIS phase is obtained by maximizing the Eve-RIS inserted deceiving channel. Our analysis and results show that even with a passive RIS, our proposed Eve-RIS can achieve a high key match rate with legitimate users and is resistant to most of the current defensive approaches. This means the novel Eve-RIS provides a new eavesdropping threat on PL-SKG, which can spur new research areas to counter adversarial RIS attacks.

5. RS2 Activity Summary

Publications: Z. Wei, B. Li, W. Guo, 'Adversarial Reconfigurable Intelligent Surface Against Physical Layer Key Generation,' IEEE Transactions on Information Forensics & Security, vol.18, Apr 2023

In the blue-teaming work, we propose an adversarial learning framework between legitimate parties (namely Alice and Bob) to address this Man-in-the-middle malicious RIS (MITM-RIS) eavesdropping. First, the theoretical mutual information gap between legitimate pairs and MITM-RIS is deduced. Then, Alice and Bob leverage generative adversarial networks (GANs) to learn to achieve a common feature surface that does not have mutual information overlap with MITM-RIS. Next, we aid signal processing interpretation of black-box neural networks by using a symbolic explainable AI (XAI) representation. These symbolic terms of dominant neurons aid feature engineering-based validation and future design of PLS common feature space. Simulation results show that our proposed GAN-based and symbolic-based PL-SKGs can achieve high key agreement rates between legitimate users and is even resistant to MITM-RIS Eve with the knowledge of legitimate feature generation (NNs or formulas). This therefore paves the way to secure wireless communications with untrusted reflective devices in future 6G.

Publications: Z. Wei, W. Hu, W. Guo, 'Explainable Adversarial Learning Framework on Physical Layer Secret Keys Combating Malicious Reconfigurable Intelligent Surface,' IEEE Transactions on Information Forensics & Security, under review, 2024.

5. RS2 Activity Summary

RS2 C: Identifying Cybersecurity Risks and Mitigation Measures for Internet of Nano/Bio Swarms

Molecular networking has been identified as a key enabling technology for Internet-of-Nano-Things (IoNT): microscopic devices that can monitor, process information, and act in a wide range of medical applications. As the research matures into prototypes, the cybersecurity challenges of molecular networking are now being researched on at both the cryptographic and physical layer level. Due to the limited computation capabilities of IoNT devices, physical layer security (PLS) is of particular interest. As PLS leverages on channel physics and physical signal attributes, the fact that molecular signals differ significantly from radio frequency signals and propagation means new signal processing methods and hardware is needed. Here, we review new vectors of attack and new methods of PLS, focusing on 3 areas:

- (1) information theoretical secrecy bounds for molecular communications,
- (2) key-less steering and decentralized key-based PLS methods, and
- (3) new methods of achieving encoding and encryption through bio-molecular compounds. The review will also include prototype demonstrations from our own lab that will inform future research and related standardization efforts.

Publications: S. Qiu et al., 'Review of Physical Layer Security in Molecular Internet of Nano-Things,' in IEEE Transactions on NanoBioscience, vol. 23, no. 1, pp. 91-100, Jan. 2024

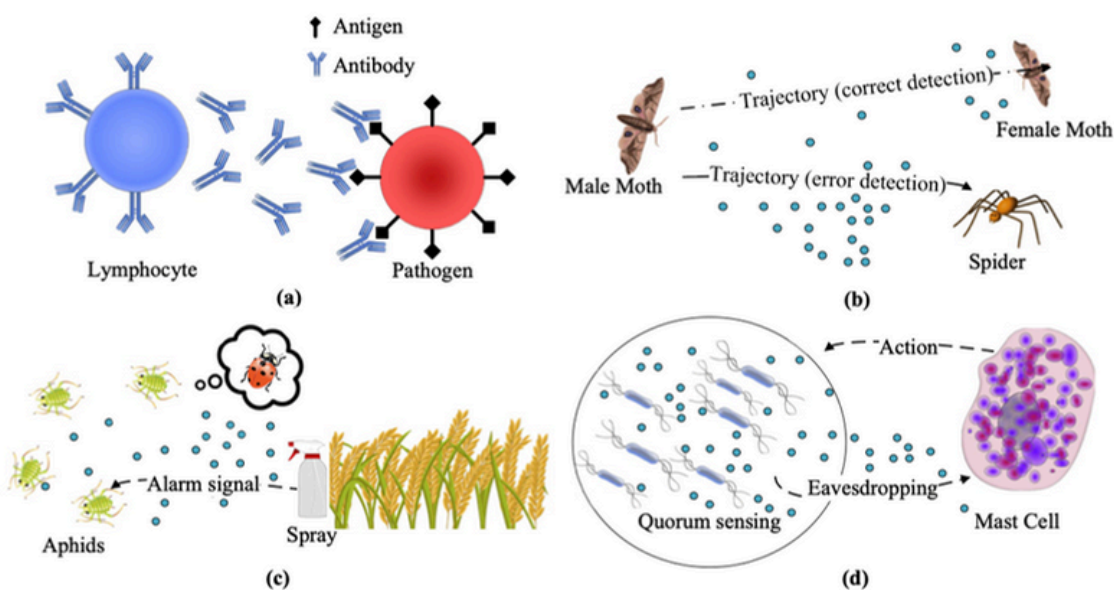


Figure 6: Natural attack vectors and cyber eavesdropping examples in networked nanobots for Internet-of-Bio-Nano-Things

5. RS2 Activity Summary

In this context, molecular communication (MC) is an emerging new communication paradigm where information is conveyed by chemical signals. It has been recognized as one of the most promising physical layer techniques for the future Internet of Bio-Nano Things (IoBNT) - Figure 2.6, which enables revolutionary applications beyond our imagination. Compared with conventional communication systems, MC typically demands a higher security level as the IoBNT is deeply associated with the biochemical process. Against this background, this article first discusses the security and privacy issues of IoBNT with MC. Then, the physical-layer countermeasures against the threat are presented from an interdisciplinary perspective concerning data science, signal processing techniques, and the biochemical properties of MC. Correspondingly, both the keyless and key-based schemes are conceived and revisited. Finally, some open research issues and future research directions for secrecy enhancement in IoBNT with MC are put forward.

Publications: Y. Huang et al., 'Physical-Layer Counterattack Strategies for the Internet of Bio-Nano Things with Molecular Communication,' in IEEE Internet of Things Magazine, vol. 6, no. 2, pp. 82-87, June 2023

Open Challenges from RS2

Assured Autonomy in Defence & National Security: Much of the TAS research has assumed benevolent, negligent, or low-end malicious actors, often operating on a relatively infrequent basis. What-if autonomy must operate with assurance in environments with persistent malicious actors that are adaptive, highly sophisticated, and highly motivated to succeed. They may seek to undermine position navigation and timing (PNT) of autonomous systems and cause tactical crashes, gamify the command and control to cause strategic mistakes, or erode societal cohesion through generative propaganda. This is particularly worrying for high value assets, where the opponent often only needs to succeed once. Here, new developments in language models may have lowered the barrier to entry for low-end adversaries, lowering the knowledge and effort required to perform a sophisticated attack. Security in autonomy that derives attributes from obscurity and complexity may no longer a viable defence, and defence for national security for assured autonomy is rapidly needed.

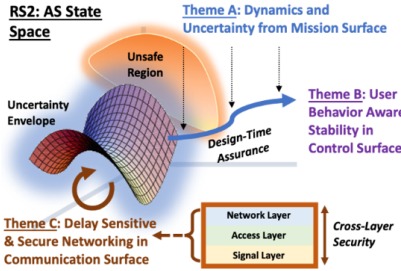
RS-2B: Securing the Control and Navigation Surfaces

Cranfield University School of Aerospace, Transport and Manufacturing

Research Assistant: Emre Saldiran, emre.saldiran@cranfield.ac.uk
 Research Assistant: Aykut Cetin, aykut.cetin@cranfield.ac.uk
 Investigator: Prof. Gokhan Inalhan, inalhan@cranfield.ac.uk



1. Role of the RS-2B

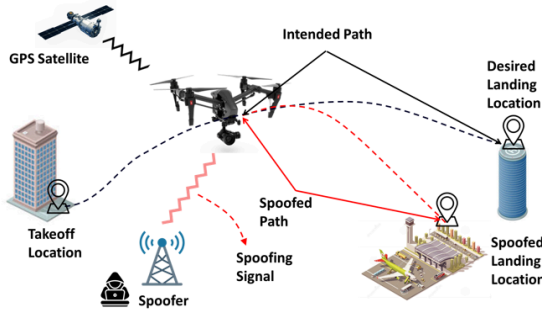


Ability of runtime adaptations of control and navigation systems over attacks or "perceived" attacks:

- Adversaries
- Environment uncertainties
- Degraded performance



2. AI-aided Visual Inertial Navigation (VIN) for GPS-denied Environments and GPS Spoofing Detection



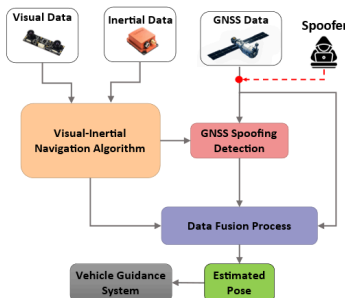
Operations in Urban Airspace

- GNSS is one of the most vulnerable system against cyber-attacks such as jamming and spoofing. These attacks are harmful and difficult to detect
- GNSS system should be supported by utilising multi-sensor pose estimation algorithms not only to detect the attacks but also to provide safety for the vehicle.

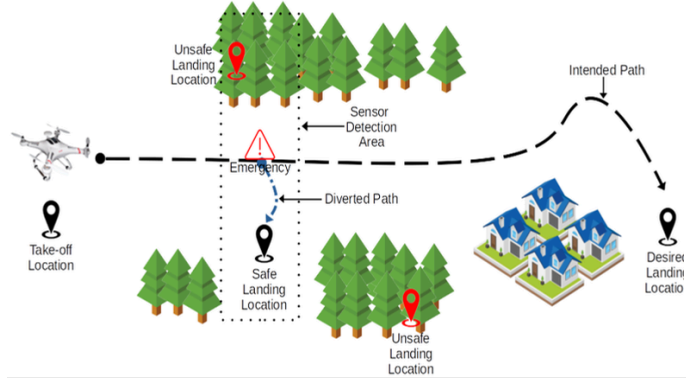
Research Proposal

- Designing AI-aided Visual-Inertial navigation system to support the GNSS in the presence of spoofing attacks.
- Combining the AI-based solutions with classical filter-based approach
- Improving pose estimation performance in austere environments

AI-aided VIN System and GPS-Spoofing Detection Overview

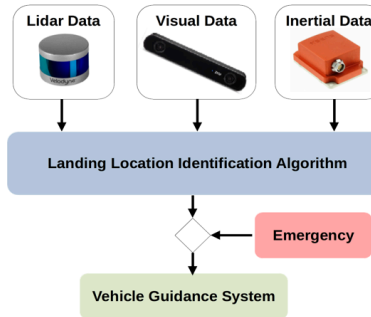


3. AI-aided Landing Location Identification for Emergency Situations and GPS-Spoofing



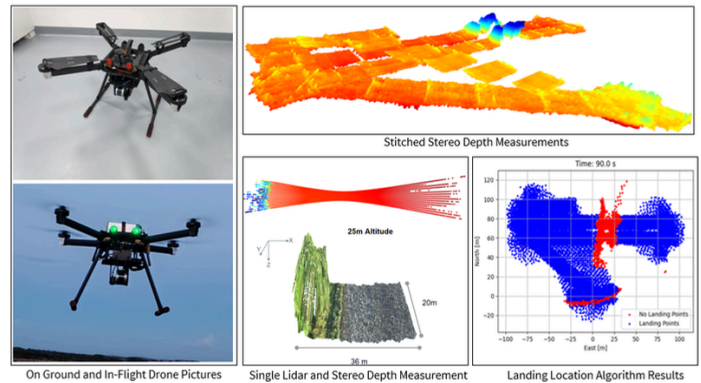
Operations in Urban and Rural Airspace

- Thousands of drone are envisioned to fly over urban and rural area with austere environments.
- They can experience multiple emergency like low battery, loss of trust and GNSS spoofing.
- Even the ones equipped with visual inertial navigation capability need to find safe landing location autonomously.



Research Proposal

- Develop an AI-aided landing location identification system for safe emergency landings.
- Support VIN system to enhance overall system trustworthiness, utilizing vision sensor capabilities.
- Improve system safety in urban and austere environments.



4. Conclusions

AI-aided Visual-Inertial Navigation System Design

We have developed AI aided visual inertial navigation algorithms providing robust navigation solution accuracy in austere environments subject to both the loss of GNSS and GNSS spoofing. The results demonstrate robust spoofing detection and autonomous navigation in challenging environments, providing a cornerstone capability towards trustworthy autonomous systems.

AI-aided Landing Location Identification for Emergency Situations and GPS-Spoofing

We have developed and demonstrated an AI-aided landing location identification system to enhance system trustworthiness, particularly supporting Visual Inertial Navigation (VIN) and improving overall safety in diverse environments. This capability is integral in urban and rural airspace operations in which reliability and safety while demonstrating a trustworthy autonomous system.



This work is supported by the Engineering and Physical Sciences Research Council [grant number: EP/V026763/1]

6. RS3 Activity Summary

Secure AS require new methodologies that social actors and organisations can deploy to critically interrogate the socio-technical and organisational processes that inform their development. To this end, the project has developed an innovative model of ‘participatory backcasting’ adapted to the particular challenges of AS whole system design.

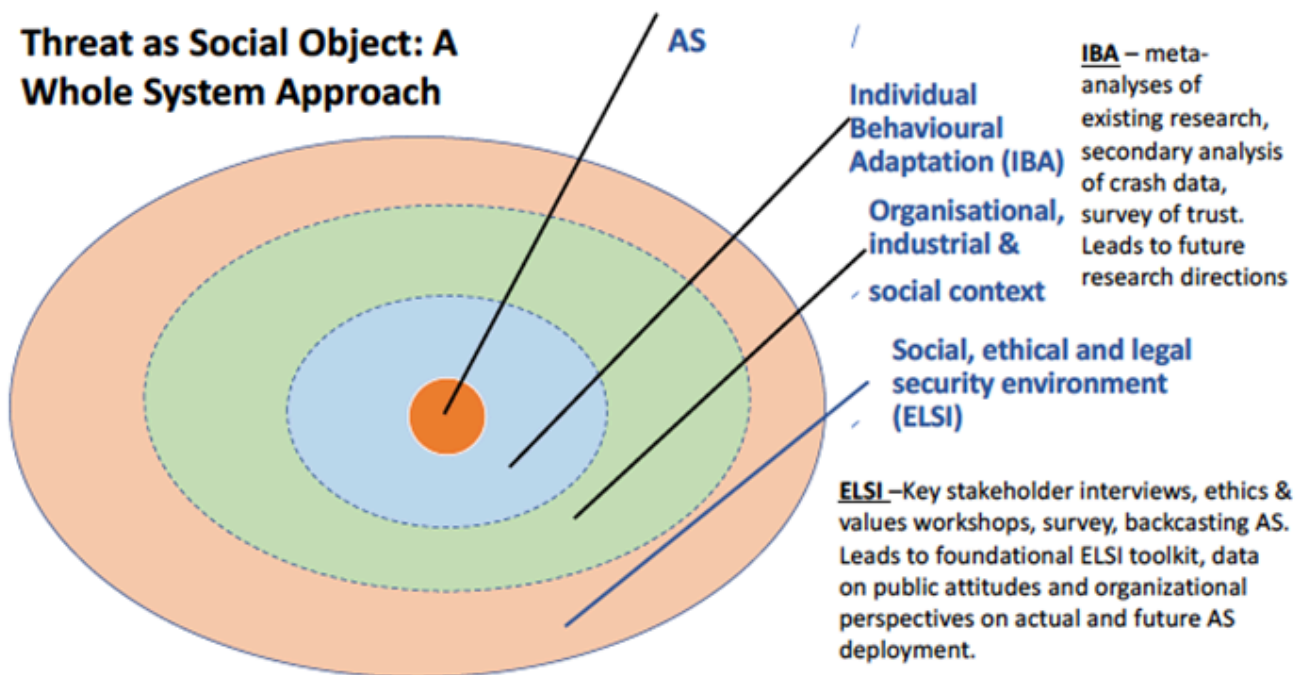


Figure 7: Threat as Social Object: Whole System Approach.

For the secure application and deployment of AS it is imperative that ethical, legal, and social issues are surveyed and understood on an ongoing basis. Both proximal users (those in direct interaction with the AS) and distal interactants (wider society) must be engaged in their design and use. TAS-S has developed a foundational toolkit to encourage this engagement. This will require continuous updating as new AS scenarios are brought to market, facilitated through crowdsourcing experience and distilling this from a security perspective.

6. RS3 Activity Summary

RS3A: Individual Behavioural Adaptation

a) Development of hypotheses about how trust in autonomous systems is built and lost to address the learning and appropriation of trust as detailed in the TAS-S proposal. A questionnaire using validated scales was distributed to over 300 participants to test several different ideas about how trust in autonomous systems develops. The influence of dispositional trust, social trust, personality, tendency to anthropomorphise, the form of the association between experience with automated systems and trust in them, and effects of age, sex, and education was investigated. Only trust in regulators and system manufacturers and experience had significant correlations with trust in autonomous systems. No other predictions were upheld. This study is under review with the Journal of Trust called 'Determinants of stated trust in automated systems; a comparative study of different predictions'.

b) A field study testing behavioural adaptation to ADAS, as described in the TAS-S proposal was conducted. Experience is a concept in human factors in automated systems (AS) research without any consensus about its definition. To investigate how a scale for canvassing experience could and should be constructed, data from a survey (1) and a field study (2) of the use of adaptive cruise control were used. In Study 1, ADAS use was strongly correlated with trust in AS and similar scales, while experience with other types of AS was not. In Study 2, predictions concerning knowledge of Adaptive Cruise Control, use of Tesla's Autopilot in the field and mood were not upheld. Only trust in AS yielded the expected result. There were strong indications that experience with AS is a multifaceted concept which need to be investigated in more detail concerning how it should be measured. The results of this study have been submitted to Ergonomics called 'Experience with automated systems; measurement issues'.

c) A meta-analysis investigated whether there are systematic differences between published studies for the safety effect of ESC. The safety effects ranged from 38 to 77 percent reduction of crashes but these effects were heterogeneous and differed depending on the methods used. Most importantly, information which could have allowed more precise analyses of the moderators were missing in most publications. Further investigations into the effects of ESC on safety using different methodologies are warranted. The paper has been submitted for publication to the Journal of Safety Research called 'Meta-analysis of the safety effect of Electronic Stability Control'.

6. RS3 Activity Summary

d) Following on from the meta-analysis, national US data for fatal crashes, driving exposure and market penetration of electronic stability control (ESC) for 1991-2021 were used to test whether ESC safety effects may be smaller due to behavioural adaptation, and increased access to ESC by less safe drivers. The analyses showed a downward trend in these crash types were generally present before ESC was introduced, and that the trends thereafter were weaker. Although some trends were consistent with effects of ESC, they were markedly smaller than the projected ones proposed by researchers. The common statement that ESC is very effective as a safety benefit is not well founded due to methodological issues in empirical research. Many predictions are suspiciously large for various automated features of vehicles. If automated vehicles were as effective as claimed, there would hardly be any crashes at all.

Impact

- a) The Role of Behavioural Science in Transport Webinars, 18-19 May 2021
- b) Lisa Dorn joined speakers and moderators from Behavioural Insight Unit, CCAV, Highways England, SWARCO, University of Leeds and JLR to discuss the role that behaviour sciences play in ensuring public trust and desirability of new technologies and how they can support deployment.
- c) Lisa Dorn served as an expert member of BS ISO 39003 Road traffic safety (RTS) to develop guidance on ethical considerations relating to safety for autonomous vehicles. The standard was published in 2023.
- d) af Wåhlberg, A. E., & Dorn, L. (submitted). Determinants of stated trust in automated systems; a comparative study of different predictions.
- e) af Wåhlberg A.E. & Dorn L (Submitted). Experience with automated systems; measurement issues.
- f) af Wåhlberg A.E. & Dorn L. (Submitted). Meta-analysis of the safety effect of Electronic Stability Control.
- g) A.E. af Wåhlberg, L. Dorn. (2024). The effects of Electronic Stability Control (ESC) on fatal crash rates in the United States, Journal of Safety Research, Volume 88, 217-229, <https://doi.org/10.1016/j.jsr.2023.11.008>.

6. RS3 Activity Summary

RS3 B&C: Ethical, Legal and Social Issues

Technical/conceptual progress

Recognising that both individual and collective behaviour adaptation are critical elements to long-term security in TAS, partnerships with civil and commercial institutions are central to understanding AS security and trust in diverse end-user environments. Evaluation of the securitized risks and longer-term societal response aim to inform future studies and legal recommendations. Within RS3B&C, the focus has been on collaboration with National Highways and Entopy to identify these risks and longer-term strategies to alleviate them. These collaborations have been hugely valuable and generated many important insights that have informed publications and reports, mentioned below. They also form the foundation of a key RS3 output, the ELSI-AS Toolkit, mentioned in 2). We are about to deliver the Transitions Report to NH, offering reflections and guidance on issues explored during our partnership, pertaining to three main themes 1) Connected and Autonomous Vehicles, 2) National Highways as a Data Manager, 3) Connected and Autonomous Plant. The second collaboration is with Entopy, a small commercial organisation developing AI platforms for real time data analytics in freight and transport across Europe. The first official partnership meeting happened in early February 2024. Interviews are currently being conducted with members of the Entopy team, and a collaborative workshop is scheduled for April. This work will feed directly into the ELSI-AS Toolkit, for which we will be working with a graphic designer to put together a set of web hosted downloaded resources for organisations to conduct their own ethical reflections and activities.

Departing from theoretical works within the humanities and complexity science, which propose that regulatory constraints amongst social agents are enacted in the everyday local interactions, Alverenga Michelin's research seeks to understand how human-machine cognitive interaction constitute and regulate the making of distributed autonomous technology. She conducted ethnographic research involving 8 participants, with a total of 26 hours of qualitative interviews and 2 hours of observation of coding practice, to draw insights into how hybrid cognitive mechanisms regulate the behaviour of the system. Preliminary results shed light into how memory plays a central role in the constitutive and regulatory processes of autonomous systems development.

6. RS3 Activity Summary

Publications:

1. Moffat L., (2024) Ethics through the Wash: Narratives of Scandal in Autonomous Systems Research, Journal of Responsible Innovation (in press)
2. Moffat, L. (2023). Relational Approaches to Autonomous Systems Ethics. In Proceedings of the First International Symposium on Trustworthy Autonomous Systems (pp. 1-7).
3. Abeywickrama, D. B., Bennaceur, A., Chance, G., Demiris, Y., Kordoni, A., Levine, M., ... & Eder, K. (2023). On specifying for trustworthiness. Communications of the ACM, 67(1), 98-109.
4. Moffat, L., May-Chahal, C., Deville, J., Dorn, L. 'Threat, Security and Safety in the Public Imaginary of Autonomous Vehicles' (in preparation)

Conference talks/Posters:

- Workshop STS Austria 'Digital Living, Digital Infrastructure'. Alverenga Michelin presentation: The Collective Cyborg Body (Sep 2021). Description: Presented previous research conducted at Goethe University and future work at Lancaster University.
- LIRA 2022 conference on Intelligent, Robotic, and Autonomous Systems (Jun 2022). Description: Organized the LIRA conference with sessions in five of LIRA research themes: Security and Defense (Alverenga Michelin invited external speaker), Nuclear Decommissioning, Fundamentals of IRAS, Society and Human Behavior, and Biomedical. Last session included a workshop to identify common interests for collaboration.
- LIRA 2022 conference on Intelligent, Robotic, and Autonomous Systems (Jun 2022). Alverenga Michelin presentation title: Distributed Cognition in Human-Machine Systems.
- Workshop Trustworthy Autonomous Systems (TAS) – Security (Nov 2022). Alverenga Michelin presentation: Distributed Epistemic Systems.
- May 2023 'AI Ethics through Design' UAL Service Design Guest Lecture, presented by Moffat, University of the Arts London
- Engagement with Industry/Government: International benchmark of cyber career policy frameworks (UKCSC/ENISA/NICE) (Jun 2023). Description: As a part-time policy and research manager of the UK Cyber Security Council, Alverenga Michelin elaborated a policy paper reporting an international benchmark between cyber career frameworks from the following: UK Cyber Security Council (the UK Cyber Career Framework), NICE (the US Workforce Framework for Cybersecurity/NIST), ENISA (the EU Cyber Security Framework).

6. RS3 Activity Summary

Conference talks/Posters (cont):

- Engagement with Industry/Government: Revision of the Cyber Security Professional Standards at UK Cyber Security Council (Aug 2023). Description: Alverenga Michelin participated in the review of the UKCSC (UK Cyber Security Council) professional standards that include 16 specialisms (as of March of 2024). Working part-time as Policy and Research Manager for the UKCSC, Juliana engaged in internal and external meetings with the Professional Standards Working Group (PSWG) to discuss the revision of the 16 specialisms currently defined by the UK government as specialized areas of knowledge and skills within the UK cyber security industry.
- STSMN (Science and Technology Studies in the Midlands and North) meeting hosted by the University of York (Sep 2023). Alverenga Michelin presentation: Distributed Epistemic Systems. The meeting relaunched the former STS4C (STS Four Cities Consortium) as STSMN (reads system). STSMN is a network that brings together STS scholars from universities in Lancaster, Leeds, Nottingham, Sheffield, and York.
- Lancaster University Sociology Dept Seminar (Nov 2023) Moffat presentation 'Dancing through the AI Ethics Machine
- Connected Places Catapult Freight Innovation Cluster Workshop (Nov 2023) Moffat, Deville & May-Chahal 'Ethics and security in autonomous systems for freight innovation'
- Response to DSIT consultation on security and resilience of UK data centres (Feb 2024). Description: Alverenga Michelin coordinated the UKCSC response to the 2024 DSIT (Department for Science, Innovation, and Technology) public consultation titled 'Protecting and enhancing the security and resilience of UK data infrastructure'. This activity involved liaising with UKCSC staff, working group members, and advisors to define the UKCSC written response to the consultation.
- March 2024 'Transport Panel' TAS Showcase

Future Plans

- ELSI Toolkit – We are currently in the early stages of designing an ELSI toolkit, a set of printable online resources that help facilitate ethical reflections by organisations either designing or working with AS. We will be securing funding to work with a graphic designer on creating a visual identity and interface for the resources.
- Entopy Collaboration – We will be drawing on our experience working with NH and adapting our methods of creative participatory workshops to help understand the challenges and opportunities facing Entopy as a commercial enterprise in the market of data analytics and cyber security. We will be conducting interviews with key members of the team to supplement these findings, which will help inform ongoing design of the ELSI toolkit, and tailored transitions guidance to be fed back to Entopy.

Trust, Experience and Behavioural Adaptation to In-Vehicle Technology

Cranfield University

Lisa Dorn and A. E. af Wåhlberg

Research Problem

- Studies reporting a safety benefit of vehicle technology do not consider the long-term impact. Safety and security may be impacted by behavioural adaptation over time.
- Trust and experience is expected to play a role in behavioural adaptation.
- Determinants of trust and the role of experience in interaction with in-vehicle technology need to be understood.
- Four key papers in press or submitted aim to investigate these research problems.

Paper 1 Determinants of Trust in Automated Systems:

- **INTRODUCTION:** Hypotheses about trust in autonomous systems (AS) are sometimes theoretical and sometimes empirical.
- **AIM:** To test competing hypotheses about how trust in AS is formed.
- **METHOD:** Predictions from AS trust literature were tested using validated scales measuring influence of dispositional trust, social trust, personality, tendency to anthropomorphise and effects of experience with AS age, sex, and education in a self-report study (N+300).
- **RESULTS:** Trust in regulators and system manufacturers and experience significantly correlated with trust in AS. No other predictions were upheld.
- **DISCUSSION:** With no personal experience, trust may be determined by social trust in the makers and/or regulators of AS technology. Trust may be sensitive to the sample selected and measurement methods used.

Paper 2: Experience with Automated Systems

- **INTRODUCTION:** Humans are adept at learning from experience and altering behaviour to better suit the environment, but we have little or no inherited understanding of AS.
- **AIM:** To develop principles for how experience with AS should be measured.
- **METHOD:** A survey study (Study 1: N= 209) and a field study (Study 2: N=22) of the use of adaptive cruise control (ACC) was conducted.

Overall Conclusions:

- Trust in AS is dependent upon experience and trust in regulators and manufacturers Safety effect of ESC is not well founded due to methodological issues
- Safety predictions are suspiciously large for various automated vehicle features If automated vehicles were as effective as claimed, there would hardly be any crashes at all
- Further research to investigate behavioural adaptation to automated vehicles is warranted
- How users adapt to AS can compromise safety and security

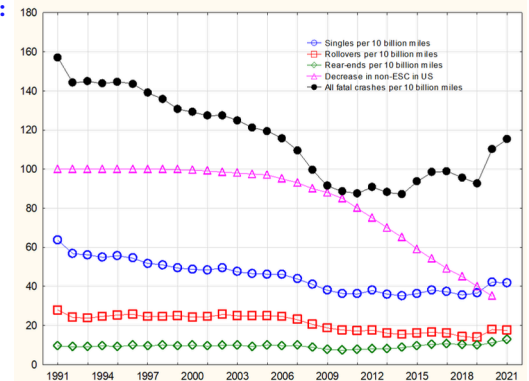
- **RESULTS:** Study 1, ADAS use strongly correlated with trust in AS and similar scales, while experience with other types of AS was not. Study 2, Trust in AS correlated with experience as in Study 1.
- **DISCUSSION:** Experience with AS is a multifaceted concept and must be fully investigated to construct a validated scale.

Paper 3: ESC Meta-analysis

- **INTRODUCTION:** Electronic Stability Control (ESC) improves manoeuvrability and claims to reduce several types of crashes.
- **AIM:** To estimate average effects of ESC on different types of crashes for all available studies and possible moderators.
- **METHOD:** A meta-analysis of ESC effects on crashes investigated whether reported effects were created by the method selected. The number of samples was increased by averaging values within samples.
- **RESULTS:** Safety effects were similar to previous meta-analyses but methodological faults in the studies could not be determined due to heterogeneity of effects, small number of studies and lack of information about certain characteristics.
- **DISCUSSION:** Methodological problems may inflate effects and these results call into question the conclusions which have calculated or discussed expected benefits of ESC at the national level.

Paper 4: Effects of ESC on Fatal Crashes

- **INTRODUCTION:** ESC empirical studies show strong effects on some crashes types but drivers choosing vehicles with ESC may be more safety-conscious.
- **AIM:** To investigate whether decreases in crash types have taken place at the same rate as ESC has increased, while other crash types would not follow this trend.
- **METHOD:** US Fatal crash data and ESC market penetration data for light four-wheel vehicles, and estimates of yearly miles travelled to follow the trends of different types of crashes over time were compared to predicted effects.
- **RESULTS:**



- **DISCUSSION:** An over-estimation of ESC effect due to self-selection and behavioural adaptation over long time periods and a diminishing effect of ESC lead to a lower-than-expected effects of ESC on safety. A similar effect could also operate with the introduction of AS.



This work is supported, in part, by the Engineering and Physical Sciences Research Council [grant number: EP/V026763/1]

Towards human centred threat assessment

Security Node Lancaster University & Cranfield University

Lancaster University: Corinne May-Chahal, Joe Deville, Catherine Easton, Luke Moffat

Cranfield University: Lisa Dorn, Anders Af Wählberg

“Those who are subject to manipulation, along with responsible governance actors, need technologies of humility the most to consider vulnerabilities, the framing and distribution of risks, and to learn together”

Monika Buscher et al (2022)

Overview

Within the TAS-S Research Strand 3 (RS3) team, we have explored ways of bringing **insights from sociology, law, and psychology** to delivering Autonomous Systems (AS) that both deliver meaningful security and are ethically responsive.

Our research has aimed to highlight the value of bringing **an expanded understandings of both ‘security’ and ‘safety’** to the design and deployment of AS, fully capable of taking account of the behavioural patterns and existential and social concerns that may be regarded as threats to AS within social systems.

A particular focus has been on understanding how questions of AS security relate to **Connected and Autonomous Vehicles (CAVs)**, including collaborations with National Highways and research into the role of behavioural adaptation within AS security.

- ❖ How can threat assessment better integrate people’s engagements with and expectations of AS?
- ❖ If we know that people adapt to AS over time, what does this mean for threat assessment?
- ❖ How can organisations themselves adapt securely to rapidly developing and changing AS?
- ❖ How can we move beyond ethics as ‘tick box’ in AS security design and deployment?
- ❖ Are standards and regulation socially responsive?

In the context of the centrality of threat modelling to cybersecurity, we propose that **threat assessment needs to expand beyond** modelling **depending on quantifiable inputs** to incorporate a wider range of threats and other forces. Domains of relevance for AS threat assessment range from particular practices and types of expertise (law and ethics), to persistent patterns of human and organisational behaviour (experience, adaptation), to inputs coming from those that might be users of, or impacted by, AS (trust). The TAS Security node has found that:

Law and ethics

- Legal frameworks need to better respond to **how AS regulations and regulators are engaged with and understood socially**, pointing to an urgent need for **more agile and innovative forms of law-making**
- Industry alone is unlikely to have the critical capacities / tools to **set standards that are socially responsive**
- Participative contextual ethics requires **other ways of knowing AS design**

Trust

- Organisations shown to need **new tools for better assessing the trust of users and wider publics** in AS
- Survey of road users indicates **low confidence about safety of CAVs**, with a **lack of trust in businesses and government**
- This is particularly important given our research shows that trust in **AS heavily shaped by ‘social trust’** in a makers’ system and/or the regulators of AS technology

Experience

- **Behaviours in relation to AS shown to be highly dependent on experience**, however we have little or no inherited understanding of AS.
- More research required on how to construct **validated scales to measure experience with AS**.
- But also need to make room for **diverse experiences with AS**, including recognising and even ‘scaffolding’ the right to refusal and to challenge the purported technological inevitability of AS

Adaptation

- **Siloing of knowledge within organisations** around AS risks inhibiting their ability to adapt to rapidly changing AS contexts
- An **urgent need for new professional roles, expertise and tools** to enable organisations to engage more critically with AS
- Evidence that **how users adapt to AS can compromise safety and security** demonstrates a pressing need for further research into the role of behavioural adaptation, including in relation to in-vehicle technology

Discussion: Implications for security, regulation and work of other TAS-S Research Strands (1 & 2)

Our work shows that **a wider set of factors must be taken into account when designing AS and assessing threats**. Approaches to threat assessment focusing solely on the interactions between a technology and a static user could miss the wider risks generated by the shifting behavioural and social environments of AS. This has **particular implications for the development of AS regulatory frameworks**.

These insights are also **being actively fed into the work of TAS-S Research Strands 1 and 2**, to further enhance both the security and response-ability of future AS design. We are also working on a **practical toolkit for organisations**, to enable them to engage with AS design and deployment more critically and creatively, in the pursuit of a more socially response approach to AS security.

Pathways to Socially Responsive AS security

Lancaster University

Joe Deville, Catherine Easton, Corinne May-Chahal, Luke Moffat

Overview

Over the past 18 months, RS3 has been working closely with National Highways, creating collaborative spaces to explore the ethical dimensions of AS security in the context of National Highways' varied roles and initiatives. Through creative workshops, public surveys and focus groups, we identified three main themes:

- Connected and Autonomous Vehicles
- National Highways as a Data Manager
- Connected and Autonomous Plant

Survey & Focus Groups: National Highways Panel

- **Conducted Oct-Nov 2022:** National Highways' Customer Panel completed a survey about their perceptions and attitudes towards the use of Autonomous Vehicles.
- **429 Panellists:** Responding across three themes: autonomous motorway building and maintenance, other self-driving vehicles, and autonomous vehicles and data sharing.
- **Headlines: High degree of cynicism** about introducing AVs on UK SRN.

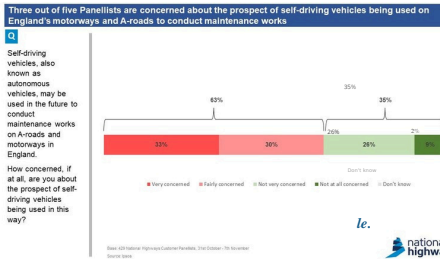


'Don't introduce autonomous vehicles until you're convinced that they are safe and secure and we have a need for them really. To introduce them just for the sake of it or for profits so wrong.'

'Who's gonna want to make sure we're all safe and secure?'

- **Low Confidence:** Concern about safety and security is high.
- **Lack of Trust:** No trusted authority in business, governmental or public sector domains to ensure CAV safety.
- **Net Opposition:** 57% oppose CAVs for both business and commercial use.
- **Lack of interest:** No significant interest in personal CAV ownership.

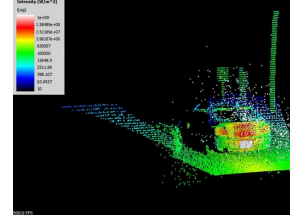
'It's all about sales for [car companies]. I don't think they're really that bothered. In fact, that I wouldn't put it past them to put things in so that you have to planned obsolescence so that you do have to keep on updating and getting repairs and things to your vehicles.'



'I think that National Highways should actually take responsibility and liability for safety on our roads [...] I think they should actually concentrate on managing and maintaining the roads we've got.'

Implications for Organisational Interaction

- **Changing landscapes:** Confronting uncertain and potentially insecure realities, making space for diverse experiences with AS.
- **Cross-sector co-operation:** Breaking down silos within and between organisations, establishing shared values.
- **New expertise:** Different kinds of professional roles needed to address new kinds of security and safety demands.
- **Building response-ability:** To manage increased scrutiny of safety, privacy, and security violations. (e.g. Mozilla privacy report).
- **Refining roles:** Balancing ideals and reals, what remits can, and should organisations reasonably adopt? Who is accountable, when, and for what?



Implications for Ethics Law and Governance

- **Participative Contextual Ethics:** Moving beyond ethics as a tick box exercise, accepting its dynamic and adaptive nature.
- **Beyond Industry Standards:** Industry cannot create standards alone, working towards socially responsive security is a collaborative endeavour.
- **Agile regulation:** Innovative forms of law-making developed from the notion that regulations are interpreted and used organisations.
- **Scaffolding refusal:** People have the right to say no, a right that is often threatened by seemingly inevitable technological change. AS must not be implemented at the expense of people's ability to opt out, or choose a different way.
- **Beyond blue sky thinking:** Ethics is not a barrier to innovation, but participative contextual ethics requires other ways of knowing AS design: what kinds of values might AS serve, beyond technical progress, entrepreneurship, and financial incentives?

Future Directions

- **ELSI Toolkit:** Interactive resources for Socially Responsive Security. Facilitating self-reflective ethical impact assessment.



Examples of a potential toolkit resource

To be used re-iteratively by organisations working with AS who are looking for creative ways to examine wider social and ethical impacts of new systems and strategies.



This work is supported, in part, by the Engineering and Physical Sciences Research Council [grant number: EP/V026763/1]

7. Case Studies: Case Study 1

Securing Drone Communications Against New Electromagnetic Meta-Surfaces and Inferring their Trustworthiness in Wider Air Spaces

Weisi Guo & Zhuangkun Wei

This case study contains 4 bodies of work – see S1.1. As drones fly close to terrestrial communication networks, their signals will interact with one of the most significant recent innovations – electromagnetic (EM) meta-surfaces / reflecting intelligent surface (RIS).

(1) These have the power to modify the EM channel, potentially expanding the channel capacity but also (we show for the first time), compromise certain security attributes such as channel-dependent key generation [S1.1].

(2) We go on to develop a world first in developing digital security attributes from physical swarm drone flight characteristics by exploiting mutual control states [S1.2]. We achieve triple verification through information theory, simulation, and experimentation.

(3) Our work in TAS-S attracted early-career-researchers (ECRs) to my wider team and two of them both won RAEng IC fellowships (Adolfo Perrusquia 2021-23; Deepak Panda 2023-25). Their work collaborated with the core TAS-S team to answer: (3a) would a third party find the drones trustworthy if they had no backdoor access to the algorithms, and (3b) how would adversarial AI attacks affect their safe navigation in future air spaces. In 3a, we developed control-physics informed machine learning to infer the intention of such drones [S1.3], with the wider body of work factoring in multiple sensory and communication aspects. In 3b, we develop robust reinforcement learning against conflict-induced spoofing in air traffic coordination [S1.4].

(4) We ask how can contextualise the RL work by using human prompts to change how RL elements (e.g., policy change or cost functions) can be affected [S1.5]. Our initial demo can turn different ethical ideas into different RL navigation outcomes.

7. Case Studies: Case Study 1

The impact of the work has been felt across a wide range of UK government and commercial bodies, many of whom were part of the original TAS-S setup, and others joined later:

- [Academic] We have since then used the outcomes to successfully develop research as part of the new £11m EPSRC 6G Future Communications Hub (led by Imperial) and won 2 RAEng IC fellowships.
- [Government] We have developed strong collaborations with drone risk at Department for Transport, given evidence to JSARC at Home Office, and have a data sharing agreement to develop trustworthy monitoring of drones across UK with NPCC.
- [Commercial] We are developing and enhancing our strategic relationships on airborne autonomy with Leonardo UK who has co-funded 2 research fellows. Other partners such as Thales UK and Saab UK have together funded 3 EPSRC PhD studentships in the above areas.
- [Public Engagement] Our work in [S1.5] is on a website (<https://ntutangyun.github.io/tas-demo/>), which allows us to diverse expert and public users to play with LLM-driven RL guided navigation.

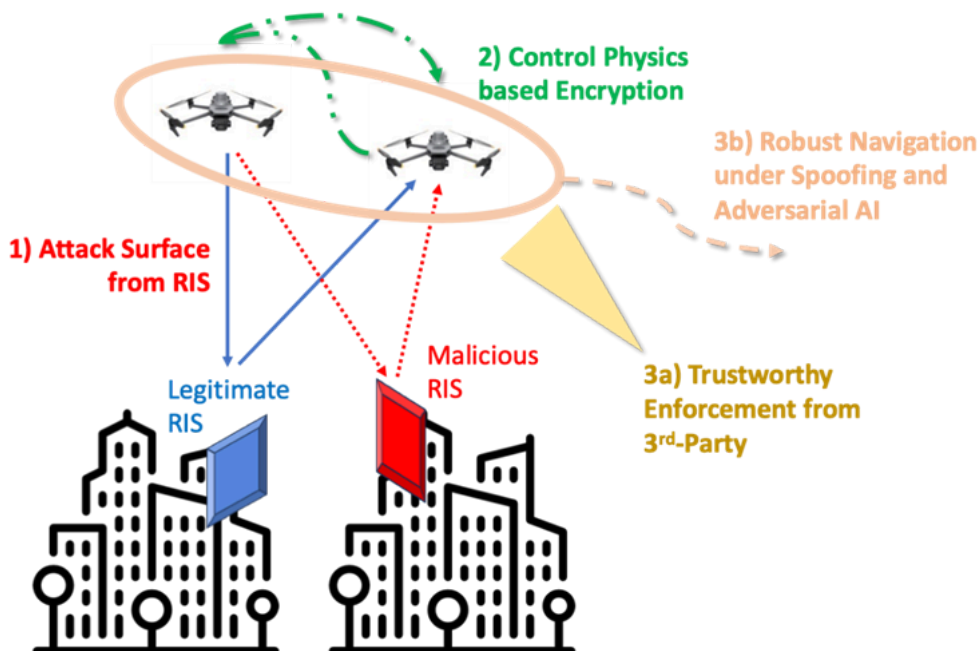


Figure 8: Swarm Drone Security and Trustworthiness: 1) communication security compromise from RIS attacks, 2) de-coupling digital communication security by exploiting swarm control physics, 3a) achieving trustworthiness from 3rd party perspective, and 3b) achieving navigation ecosystem security against adversarial AI.

7. Case Studies: Case Study 1

References

[S1.1] ‘Adversarial Reconfigurable Intelligent Surface Against Physical Layer Key Generation,’ Z. Wei, B. Li, W. Guo, IEEE Transactions on Information Forensics & Security, 2023

[S1.2] ‘Control Layer Security: Exploiting Unobservable Cooperative States of Autonomous Systems for Secret Key Generation,’ Z. Wei, W. Guo, IEEE Transactions on Mobile Computing, 2024

[S1.3] ‘Uncovering Drone Intentions using Control Physics Informed Machine Learning,’ A. Perrusquia, W. Guo, B. Fraser, Z. Wei, Nature Communications Engineering, 2024

[S1.4] ‘Action Robust Reinforcement Learning for Air Mobility Deconfliction against Conflict Induced Spoofing,’ W. Guo, D. Panda, IEEE Transactions on Intelligent Transportation Systems, 2024

[S1.5] ‘Encoding Social & Ethical Values in Autonomous Navigation: Philosophies Behind an Interactive Online Demonstration,’ Y. Tang, L. Moffat, W. Guo, C. May-Chahal, J. Deville, A. Tsourdos, ACM International Symposium on Trustworthy Autonomous Systems (TAS), Sep 2024

7. Case Studies: Case Study 2

Bringing ELSI Principles and Creative Methods to National Highways

Joe Deville & Luke Moffat

A key part of the work of the TAS-S RS3 team involved a collaboration with National Highways (NH) (formerly Highways England), to explore their potential futures with autonomous systems. This collaboration took place over the course of a year and focused in particular on engaging NH colleagues in discussions around the challenges of delivering ethically responsive and more secure futures with autonomous systems.

Changing futures with autonomous systems

For NH, autonomous systems present a range of potential future uncertainties. With a responsibility for large parts of the UK's road network, one question the organization is having to consider how it should respond to a future that could include an increasing number of Connected and Autonomous Vehicles (CAVs) on the roads. But there are other uncertainties linked to NH's possible futures with autonomous systems. For example, could road building and maintenance benefit from the increased deployment of yet-to-be-developed autonomous systems? How could autonomous systems change the way in which traffic flows is managed?

Such uncertainties open up potential ethical and security challenges. As ever more autonomous systems are deployed in and around a large and complex road network, what new vulnerabilities emerge from the combination of an increasing range of connected digital systems? How might the actual and perceived safety and security of new autonomous technologies affect confidence in the range of organizations responsible for UK's roads? Who might benefit from these changes? Who might be harmed? What unintended consequences could there be associated with such systems?

Unpacking autonomous systems with creative, ELSI methods

Members of the TAS-S RS3 team used a range of methods – combining insights from sociology and socio-legal studies – to collaboratively engage NH colleagues in such questions. One aim was to provide NH with new frameworks for discussing such questions internally, as well as to support NH in being better able to anticipate and prepare for these possible new futures with autonomous systems. This work included three workshops with NH colleagues, 1:1 in depth interviews with 8 NH colleagues with particular responsibility for considering the place of autonomous systems within NH's future, and wider scoping work with members of the public who might be affected by autonomous systems operating in and around the UK's road network. This included a survey (400+ responses) and 4 focus groups.

7. Case Studies: Case Study 2

The team’s work was underpinned by an ‘ELSI’ or ‘Ethical, Legal and Social Issues’ approach. This approach is often used by social scientific researchers interested in understanding how technologies interact with their contexts in diverse ways. RS3 researchers used it to unpack the Ethical, Legal and Social Issues connected to predicted, anticipated, and imagined changes with autonomous systems on many of the UK’s roads and the consequences of this for AS security.

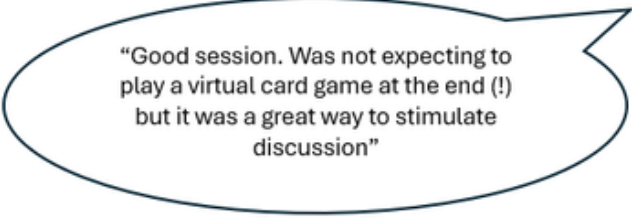
A distinctive element of the team’s work was the combination of an ELSI approach with creative methods. In the first workshop, the research team involved NH colleagues in visioning exercises, supported by online collaboration tools, to shed light on the diverse spaces where autonomous systems could have impacts on NH’s current and future areas of responsibility, and where new security challenges might arise. This was followed by a workshop focused on ethics, supported by a virtual card game which invited members of the team to collaborate on establishing agreed, priority ethical principles to potentially inform their work. A final workshop used the results from the first two workshops to inform a ‘backcasting’ exercise. The workshop involved NH colleagues in not just predicting possible futures for NH alongside autonomous systems, but also in establishing what NH’s desired futures with autonomous systems could be, and what steps would be required to begin to practically accomplish such futures.



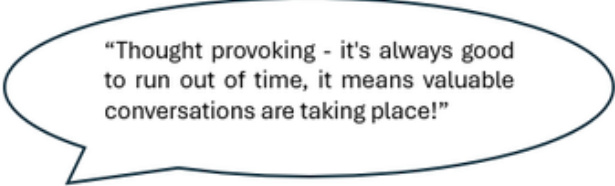
Figure 9: Example cards from game used in workshop with NH colleagues.

Feedback from NH colleagues underscored the value of such approaches (see below), with one colleague also noting during the course of a workshop how the methods being used had provided an opportunity to think through questions that usually are inhibited by organizational silos. The workshops also highlighted to participants how NH’s role might change, with an increasing dependence on autonomous systems: from a road builder and maintainer to a data manager. This, colleagues suggested, could pose new legal, ethical and security challenges in need of further attention.

7. Case Studies: Case Study 2



“Good session. Was not expecting to play a virtual card game at the end (!) but it was a great way to stimulate discussion”



“Thought provoking - it's always good to run out of time, it means valuable conversations are taking place!”

Feedback from NH colleagues on RS3 workshops

The work with wider publics, meanwhile, revealed some of the issues at stake in the context of such uncertain, technologically-mediated futures. The survey indicated just how low confidence in CAVs on UK roads is, with only around one in five of respondents in support. Around two thirds of respondents also expressed concerns about the potential for autonomous road building and maintenance technologies to be used on the UK's roads. Another revealing finding was that by far the largest range of concerns were not about specific technologies, but about the use and impact of these technologies – concerns we group together as ‘socially motivated threats’. These include worries about how others might use such technologies (28%), worries about loss of attention (12%) and worries about how autonomous systems might impact job security (9%). Such evidence points to the need – not just for NH, but for many stakeholders – to recognise that discussions about autonomous systems security should include a wider understanding of threat, when considering what consequences could stem from their development and deployment.

Diverse outputs

Over the course of the project, the team have delivered a wide range of outputs, aimed a diverse audience. Alongside conventional academic outputs, this includes short [reports](#), a [magazine piece](#) and a [video](#). The team are also in the process of finalising a toolkit, aimed at enabling organisations to use similar methods to those trialled with NH, to creatively explore both the opportunities and challenges for ethics and security posed by a wide range of new, autonomous systems.

7. Case Studies: Case Study 3

Road Traffic Safety (RTS) - Guidance on Ethical Considerations Relating to Safety for Autonomous Vehicles: ISO 39003:2023

Lisa Dorn

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Dr Lisa Dorn, TAS-S Co-I based at Cranfield University was invited to join the Technical Committee ISO/TC 241 Road Traffic Safety Management Systems in recognition of her research in the field. The committee aimed to prepare a new standard called 'Guidance on ethical considerations relating to safety for autonomous vehicles' (AVs). Several standards are available, or under development, to address the engineering and technological aspects of AVs, but none that cover the decision-making processes of the vehicle control system. The new standard called ISO39003 is from the 'family' of Standards with ISO39001 being the first. ISO39001 is a full management system standard for Road Traffic Safety Management Systems providing a framework for a safe system approach for organisations. ISO39002 gives guidance on safe commuting practices. The intent for ISO39003 is that all manufacturers and designers should consider the identified safety aspects using a self-developed, standardized, methodology for AVs. ISO39003 is applicable to vehicles in level 5 mode according to SAE J3016 in 2022, as part of its report.

The ISO39003 standard aimed to address:

- Methodology for Identification and Evaluation of aspects
- Those aspects that should be considered
- Possible outcomes of those decisions

7. Case Studies: Case Study 3

Working draft 1 to 9 of the draft ISO39003 standard was developed between October 2019 to January 2022. The Committee draft was published on February 14th, 2022. The working draft was approved further to a review and ballot by members of ISO TC 241 and the draft International Standard was published on July 11, 2022. The entire ISO approved the draft further to a review and ballot and the Final Draft International Standard ballot was initiated. BSI published ISO 39003 on 27th July 2023 to give guidance on ethical considerations with regards to road traffic safety of AVs. The ISO39003 standard does not apply to the technical method used to control the decision-making process, nor does it give any guidance on the desired outcomes of those decisions; it gives guidance on ethical aspects for consideration in the design of the decision-making process. The standard does not offer the technical precision to prescribe the required controls but offers a set of 'protocol guidelines' that all decision makers for the design of AVs could choose to self-certify against to assure that the desired necessary ethical considerations were addressed during design and effectively controlled. ISO39003 considers those aspects of AVs that require considerations to be made by the designer/manufacturer to ensure that key ethical aspects are not overlooked or disregarded. It provides manufacturers and distributors with a mechanism to enable formal declaration of compliance to an International Standard and will give assurance to purchasers, end-users, and society, that the vehicles' design has considered and addressed the ethical issues identified within the standard.

BS ISO 39003:2023

[Road traffic safety \(RTS\). Guidance on ethical considerations relating to safety for autonomous vehicles](#)

8. Publications, Media & Products

Publications 2024

Af Wåhlberg, A.E. and Dorn, L. (2024) 'Meta-analysis of the safety effect of electronic stability control', *Journal of safety research*, 90, pp. 350–370. Available at: <https://doi.org/10.1016/j.jsr.2024.07.004>.

Af Wåhlberg, A.E. and Dorn, L. (2024) 'The effects of Electronic Stability Control (ESC) on fatal crash rates in the United States', *Journal of safety research*, 88, pp. 217–229. Available at: <https://doi.org/10.1016/j.jsr.2023.11.008>.

Chuah, E. and Suri, N. (2024) 'An empirical study of reflection attacks using NetFlow data', *Cybersecurity*, 7(1), pp. 13–22. Available at: <https://doi.org/10.1186/s42400-023-00203-7>.

Deville, J. (2024) *The Future of Roads* (video). Lancaster University Management School. Lancaster University. Available at: [The Future of Roads - YouTube](https://www.youtube.com/watch?v=...).

Gunasekera, O.V.W., Sogokon, A., Gouglidis, A., Suri, N. (2024) 'Real Arithmetic in TLAPM', in: Benz, N., Gopinath, D., Shi, N. (eds) *NASA Formal Methods. NFM 2024. Lecture Notes in Computer Science*, vol 14627. Springer, Cham. Available at: https://doi.org/10.1007/978-3-031-60698-4_8.

Guo, W. et al. (2024) 'Control Layer Security: A New Security Paradigm for Cooperative Autonomous Systems', *IEEE vehicular technology magazine*. IEEE, pp. 93–102. Available at: <https://doi.org/10.1109/MVT.2023.3290773>.

Kastanakis, S. et al. (2024) 'Investigating Location-aware Advertisements in Anycast IP Networks', in *Proceedings of the 2024 Applied Networking Research Workshop*. New York, NY, USA: ACM, pp. 15–22. Available at: <https://doi.org/10.1145/3673422.3674885>.

Li, Y. et al. (2024) 'Federated Adversarial Learning for Robust Autonomous Landing Runway Detection'. *Artificial Neural Networks and Machine Learning – ICANN 2024*, Lugano, Switzerland, 2024, *Proceedings, Part VI*. (pp. 159–173). Available at: <https://doi.org/10.48550/arxiv.2406.15925>.

Li, Y., Angelov, P. and Suri, N. (2024) 'Rethinking Self-supervised Learning for Cross-domain Adversarial Sample Recovery', in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–7. Available at: <https://doi.org/10.1109/IJCNN60899.2024.10650687>.

8. Publications, Media & Products

Publications 2024

May-Chahal C, Deville J, Moffat L, Guo W, Tang Y, Tsourdos A (2024) 'Encoding Social & Ethical Values in Autonomous Navigation: Philosophies Behind an Interactive Online Demonstration', in Proceedings of the Second International Symposium on Trustworthy Autonomous Systems (TAS '24). Association for Computing Machinery, New York, NY, USA, Article 21, 1–9. Available at: <https://doi.org/10.1145/3686038.3686044>.

Moffat L. (2024). Ethics through the Wash: Narratives of Scandal in Autonomous Systems Research. Journal of Responsible Innovation Special Issue (forthcoming)
Pellicer, A.L. et al. (2024) 'UNICAD: A Unified Approach for Attack Detection, Noise Reduction and Novel Class Identification', in 2024 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8. Available at: <https://doi.org/10.1109/IJCNN60899.2024.10651159>.

Perrusquía A, Guo W, Fraser B, Wei Z. (2024) 'Uncovering drone intentions using control physics informed machine learning', Communications engineering, 3(1), pp. 36–14. Available at: <https://doi.org/10.1038/s44172-024-00179-3>.

Perrusquia, A. and Guo, W. (2024) 'Reservoir Computing for Drone Trajectory Intent Prediction: A Physics Informed Approach', IEEE transactions on cybernetics, 54(9), pp. 1–10. Available at: <https://doi.org/10.1109/TCYB.2024.3379381>.

Perrusquia, A. and Guo, W. (2024) 'Trajectory Inference of Unknown Linear Systems Based on Partial States Measurements', IEEE transactions on systems, man, and cybernetics. Systems, 54(4), pp. 2276–2286. Available at: <https://doi.org/10.1109/TSMC.2023.3344017>.

Qiu S., Wei Z., Huang Y., Abbaszadeh M., Charmet J., Li B., Guo W. (2024) Review of Physical Layer Security in Molecular Internet of Nano-Things, IEEE transactions on nanobioscience. United States: IEEE, pp. 1–1. Available at: <https://doi.org/10.1109/TNB.2023.3285973>.

Wei, Z. and Guo, W. (2024) 'Control Layer Security: Exploiting Unobservable Cooperative States of Autonomous Systems for Secret Key Generation', IEEE Transactions on Mobile Computing. IEEE, pp. 1–12. Available at: <https://doi.org/10.1109/TMC.2024.3369754>.

8. Publications, Media & Products

Publications 2024

Y. Li, P. Angelov, and N. Suri (2024) 'Self-supervised representation learning for adversarial attack detection', in proceedings of European Conference on Computer Vision (ECCV). Available at: <https://doi.org/10.48550/arxiv.2407.04382>.

Y. Wang, Y. Lu, Z. Dong and Y. Dong (2024) 'Privacy-preserving decentralised federated learning for short-term load forecasting', 43rd Chinese Control Conference (CCC), Kunming, China, 2024, pp. 9046-9051. Available at: <https://doi.org/10.23919/CCC63176.2024.10662194>.

Yang, J. et al. (2024) 'Distributed Neighbor Selection for Second-order Semi-Autonomous Networks', in 2024 American Control Conference (ACC), pp. 2612-2617. Available at: <https://doi.org/10.23919/ACC60939.2024.10644956>.

8. Publications, Media & Products

Publications 2023

Abeywickrama, D.B. et al. (2023) 'On Specifying for Trustworthiness'. Available at: <https://doi.org/10.48550/arxiv.2206.11421>.

Bildik E, Yuksek B, Tsourdos A, Inalhan G. (2023) 'Development of Active Decoy Guidance Policy by Utilising Multi-Agent Reinforcement Learning', AIAA 2023-2668. Available at: <https://doi.org/10.2514/6.2023-2668>.

C. Li and W. Guo, 'Soft Body Pose-Invariant Evasion Attacks against Deep Learning Human Detection,' 2023 IEEE Ninth International Conference on Big Data Computing Service and Applications (BigDataService), Athens, Greece, 2023, pp. 155-156. Available at: <https://doi.org/10.1109/BigDataService58306.2023.00032>.

Candan, B. et al. (2023) 'Intelligent Wargaming Approach to Increase Course of Action Effectiveness in Military Operations'. Available at: <https://doi.org/10.2514/6.2023-2531>.

Huang, Y. et al. (2023) 'Physical-Layer Counterattack Strategies for the Internet of Bio-Nano Things with Molecular Communication', IEEE internet of things magazine. New York: IEEE, pp. 82-87. Available at: <https://doi.org/10.1109/IOTM.001.2300029>.

Karali, H. et al. (2023) 'Data-driven Synthetic Air Data Estimation System Development for a Fighter Aircraft', in 2023 AIAA Aviation and Aeronautics Forum and Exposition (AIAA AVIATION Forum), Paper number AIAA 2023-3439. Available at: <https://doi.org/10.2514/6.2023-3439>.

Kastanakis, S. et al. (2023) 'Replication: 20 Years of Inferring Interdomain Routing Policies', in Proceedings of the 2023 ACM on Internet Measurement Conference. New York, NY, USA: ACM, pp. 16-29. Available at: <https://doi.org/10.1145/3618257.3624799>.

Li, C. et al. (2023) 'Scarce data driven deep learning of drones via generalized data distribution space', Neural computing & applications, 35(20), pp. 15095-15108. Available at: <https://doi.org/10.1007/s00521-023-08522-z>.

Li, Y., Angelov, P. and Suri, N. (2023) 'Domain Generalization and Feature Fusion for Cross-domain Imperceptible Adversarial Attack Detection', in 2023 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1-8. Available at: <https://doi.org/10.1109/IJCNN54540.2023.10191267>.

8. Publications, Media & Products

Publications 2023

Li, Y., Angelov, P. and Suri, N. (2023) 'Fuzzy Detectors Against Adversarial Attacks', in 2023 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, pp. 306–311. Available at: <https://doi.org/10.1109/SSCI52147.2023.10372061>.

Lou, J. et al. (2024) 'Real-Time On-the-Fly Motion Planning for Urban Air Mobility via Updating Tree Data of Sampling-Based Algorithms Using Neural Network Inference', Aerospace, 11(1), pp. 99-. Available at: <https://doi.org/10.3390/aerospace11010099>.

Lu, Y., Yu, Z. and Suri, N. (2023) 'Privacy-preserving Decentralized Federated Learning over Time-varying Communication Graph', ACM transactions on privacy and security, 26(3), pp. 1–39. Available at: <https://doi.org/10.1145/3591354>.

Moffat, L. (2023) 'Relational Approaches to Autonomous Systems Ethics', in Proceedings of the First International Symposium on Trustworthy Autonomous Systems. New York, NY, USA: ACM, pp. 1–7. Available at: <https://doi.org/10.1145/3597512.3600201>.

Sogokon, A. et al. (2023) 'Specifying Autonomous System Behaviour'. Available at: <https://doi.org/10.48550/arxiv.2302.10087>.

Varadharajan, V. and Suri, N. (2024) 'Security challenges when space merges with cyberspace', Space policy, 67, pp. 101600-. Available at: <https://doi.org/10.1016/j.spacepol.2023.101600>.

Wei, Z., Li, B. and Guo, W. (2023) 'Adversarial Reconfigurable Intelligent Surface Against Physical Layer Key Generation', IEEE transactions on information forensics and security, 18, pp. 2368–2381. Available at: <https://doi.org/10.1109/TIFS.2023.3266705>.

Yu, Z., Lu, Y. and Suri, N. (2023) 'RAFL: A Robust and Adaptive Federated Meta-Learning Framework Against Adversaries', in 2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS). IEEE, pp. 496–504. Available at: <https://doi.org/10.1109/MASS58611.2023.00068>.

Yukse B, Yu Z, Suri N, Inalhan G. (2023) 'Federated Meta Learning for Visual Navigation in GPS-denied Urban Airspace', in 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC). IEEE, pp. 1–7. Available at: <https://doi.org/10.1109/DASC58513.2023.10311195>.

8. Publications, Media & Products

Publications 2022

Bildik, E. et al. (2022) 'Development of Reinforcement Learning Based Mission Planning Method for Active Off-board Decoys on Naval Platforms'. In AIAA SCITECH 2022 Forum, p. 2104. Available at: <https://doi.org/10.2514/6.2022-2104>.

Chen, Y. et al. (2022) 'SlowCoach: Mutating Code to Simulate Performance Bugs', in 2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE). IEEE, pp. 274–285. Available at: <https://doi.org/10.1109/ISSRE55969.2022.00035>.

Chen, Y., Bradbury, M. and Suri, N. (2022) 'Towards Effective Performance Fuzzing', in 2022 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Charlotte, NC, USA, pp. 128-129. Available at: <https://doi.org/10.1109/ISSREW55968.2022.00055>.

Chuah, E. et al. (2022) 'A Survey of Log-Correlation Tools for Failure Diagnosis and Prediction in Cluster Systems', IEEE access, 10, pp. 1–1. Available at: <https://doi.org/10.1109/ACCESS.2022.3231454>.

Gonzalez V, O.J. and Tsourdos, A. (2022) 'A Laguerre-based Distributed Nonlinear Model Predictive Control Scheme for Dynamic Obstacle Avoidance on Multi-Rotor UAVs', in 2022 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, pp. 1632–1637. Available at: <https://doi.org/10.1109/ICUAS54217.2022.9836049>.

Gonzalez Villarreal, O.J., Rossiter, J.A. and Tsourdos, A. (2022) 'An efficient condensing algorithm for fast closed loop dual mode nonlinear model predictive control', in IET control theory & applications, 16(9), pp. 872–888. Available at: <https://doi.org/10.1049/cth2.12274>.

Hackett, W. et al. (2023) 'PINCH: An Adversarial Extraction Attack Framework for Deep Learning Models'. Available at: <https://doi.org/10.48550/arxiv.2209.06300>.

Kastanakis, S., Giotsas, V. and Suri, N. (2022) 'Understanding the confounding factors of inter-domain routing modeling', in Proceedings of the 22nd ACM Internet Measurement Conference. New York, NY, USA: ACM, pp. 758–759. Available at: <https://doi.org/10.1145/3517745.3563025>.

Luján Escalante, M.A., Moffat, L., and Büscher, M. (2022) 'Ethics through design' in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), DRS2022: Bilbao, 25 June - 3 July, Bilbao, Spain. Available at: <https://doi.org/10.21606/drs.2022.400>.

8. Publications, Media & Products

Publications 2022

Manzoor, S. et al. (2022) 'Poster: Effectiveness of Moving Target Defense Techniques to Disrupt Attacks in the Cloud', in Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. New York, NY, USA: ACM, pp. 3415–3417. Available at: <https://doi.org/10.1145/3548606.3563514>.

Manzoor, S. et al. (2022) 'Poster: Multi-Layer Threat Analysis of the Cloud', in Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. New York, NY, USA: ACM, pp. 3419–3421. Available at: <https://doi.org/10.1145/3548606.3563515>.

Manzoor, S. et al. (2022) 'ThreatPro: Multi-Layer Threat Analysis in the Cloud'. Available at: <https://doi.org/10.48550/arxiv.2209.14795>.

Soares, E., Angelov, P., Suri, N. (2022) 'Similarity-based Deep Neural Network to Detect Imperceptible Adversarial Attacks', 2022 IEEE Symposium Series on Computational Intelligence (SSCI), Singapore, Singapore, 2022, pp. 1028-1035. Available at: <https://doi.org/10.1109/ssci51031.2022.10022016>.

Wei Z, Wang L, Guo W. (2022) 'Secret Key Rate Upper-bound for Reconfigurable Intelligent Surface-combined System under Spoofing', in 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall), London, United Kingdom, pp. 1-6. Available at: <https://doi.org/10.1109/VTC2022-Fall57202.2022.10012819>.

Wei, Z., Guo, W. and Li, B. (2022) 'A Multi-Eavesdropper Scheme Against RIS Secured LoS-Dominated Channel', IEEE communications letters, 26(6), pp. 1221–1225. Available at: <https://doi.org/10.1109/LCOMM.2022.3166239>.

Wei, Z., Wang L. and Guo W. (2022) 'Secret Key Rate Upper-bound for Reconfigurable Intelligent Surface-combined System under Spoofing', in 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall), London, United Kingdom, 2022, pp. 1-6. Available at: <https://doi.org/10.1109/vtc2022-fall57202.2022.10012819>.

Yu, Z., Lu, Y., Angelov, P. and Suri, N. (2022) 'PPFM: An adaptive and hierarchical peer-to-peer federated meta-learning framework', in 2022 18th International Conference on Mobility, Sensing and Networking (MSN), Guangzhou, China, 2022, pp. 502-509. Available at: <https://doi.org/10.1109/MSN57253.2022.00086>.

Yukse, B. and Inalhan, G. (2022) 'Transition Flight Control System Design for Fixed-Wing VTOL UAV: A Reinforcement Learning Approach', in proceedings of AIAA SCITECH 2022 Forum, Paper number AIAA 2022-0879. Available at: <https://doi.org/10.2514/6.2022-0879>.

8. Publications, Media & Products

Publications 2021

CChuah, E. et al. (2021) 'Challenges in Identifying Network Attacks Using Netflow Data', in 2021 IEEE 20th International Symposium on Network Computing and Applications (NCA). IEEE, pp. 1-10. Available at: <https://doi.org/10.1109/NCA53618.2021.9685305>.

ChuahM, E., et al. (2021) 'Failure Diagnosis for Cluster Systems using Partial Correlations', in 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking, pp. 1091-1101. Available at: <https://doi.org/10.1109/ispa-bdcloud-socialcom-sustaincom52081.2021.00151>.

Lou, J., Yuksek, B., Inalhan, G. and Tsourdos, A. (2021) 'An RRT Based Method for Dynamic Mission Balancing for Urban Air Mobility Under Uncertain Operational Conditions', in 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC). IEEE, pp. 1-10. Available at: <https://doi.org/10.1109/DASC52595.2021.9594424>.

Luján Escalante, M., Moffat, L., Harrison, L., and Kuh, V. (2021) Dancing with the Troubles of AI, in Leitão, R.M., Men, I., Noel, L-A., Lima, J., Meninato, T. (eds.), Pivot 2021: Dismantling/Reassembling, 22-23 July, Toronto, Canada. Available at: <https://doi.org/10.21606/pluriversal.2021.0037>.

Soares E. and Angelov P. (2021) 'Radnn: Robust to imperceptible adversarial attacks deep neural network', TechRxiv, October 04, 2021. Available at: <https://doi.org/10.36227/techrxiv.16709359.v1>.

Yukse, Burak & Demirezen, M. & Inalhan, Gokhan & Tsourdos, Antonios. (2021). Cooperative Planning for an Unmanned Combat Aerial Vehicle Fleet Using Reinforcement Learning. Journal of Aerospace Information Systems,18(10), pp.739-750. Available at: <https://doi.org/10.2514/1.i010961>.

8. Publications, Media & Products

Interviews & Podcasts

Mirage (2021) 'Driverless cars step closer to our roads with new self-learning AI technology', 16 Dec, Available at: <https://www.miragenews.com/driverless-cars-step-closer-to-our-roads-with-695937/>

Corinne May-Chahal and Neeraj Suri (2022) 'How universities and businesses can work together to foster innovation', Barclays Eagle Lab 09 May, [Interview]. Available at: <https://labs.uk.barclays/learning-and-insights/news-and-insights/technology/how-universities-and-businesses-can-work-together-to-foster-innovation/>

Lisa Dorn (2023) Let's Talk Fleet Risk: 'Commercial Drivers – driver behaviour and improving driver coping strategies', 12 March [Podcast]. Available at: <https://www.drivingforbetterbusiness.com/podcast/episode/commercial-drivers-driver-behaviour-and-improving-driver-coping-strategies/>

Daniel Prince (2022) Cyberwire podcast: CyberWire-X 'HEAT: Episode 26: Examining the next-class of browser-based attacks', 6 March [Podcast]. Available at: [HEAT: Examining the next-class of browser-based attacks.](#)

Sean Riley (2021) Living with AI Podcast: Season 1, Episode 10: 'Challenges of Living with Artificial Intelligence', Prof Professor Neeraj Suri. (24 March) Available at: <https://www.buzzsprout.com/1447474/8022176>

Social Media

TAS-S Website. Available at: [UKRI TAS-S: Trustworthy Autonomous Systems Node in Security](#)

X/Twitter account [@TAS_Security](#)

8. Publications, Media & Products

Products

Artistic or Creative Products

Artefact (including digital) - ELSI Toolkit: A co-created physical & digital resource for organisations to conduct ethical impact self-assessment for the design and/or adoption of new autonomous technologies. This toolkit is designed for independent facilitation of workshop and reflective activities, which allow organisations working with autonomous technologies to critically assess the ethical and security impacts of their work. This will help create space for complex conversations about wider effects of autonomous technologies of organisations and society, including users, publics, and environments. The toolkit aims to provide simple and accessible resources for organisations to use to host 1 – 3 internal workshops with staff around relevant themes. Each workshop includes a facilitation guide, supported by resources. (forthcoming)

Software & Technical Products

Webtool/Application: Encoding Social Values in Autonomous Navigation: An Interactive Online Demonstration This website contains five demos, Baseline Scenario, Encoding Overview, Demo 1, Demo 2, Demo 3, and Demo 4, presenting how autonomous system developers encode ELSI (ethical, legal and social impact) principles into the decision-making process of autonomous systems. Available at: <https://ntutangyun.github.io/tas-demo/>.

Software: Uncovering Drone Intentions using Control Physics Informed Machine Learning: data This repository provides the data and code of the paper 'Uncovering Drone Intentions using Control Physics Informed Machine Learning'. Available at: https://cord.cranfield.ac.uk/articles/software/Uncovering_Drone_Intentions_using_Control_Physics_Informed_Machine_Learning_data/25204409.

8. Publications, Media & Products

Blogs

Deville, J. and Moffat, L. (2022) 'Values and Visions: TAS-S and National Highways Workshop #2', TAS-S Blog, 24 May. Available at: [Values and Visions: TAS-S and National Highways Workshop #2 - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Foster, P. (2021) 'Meeting the TAS-S team!', TAS-S Blog, 9 December. Available at: [Meeting the TAS-S team! - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Foster, P. (2022) 'The University with an Airport!', TAS-S Blog, 24 May. Available at: [The University with an Airport! - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Foster, P. (2023) 'Collaboration, collaboration, collaboration', TAS-S Blog, 10 February. Available at: [Collaboration, collaboration, collaboration - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Foster, P. (2023) 'TAS-S ESG: Organisational Challenges', TAS-S Blog, 23 May. Available at: [TAS-S ESG: Organisational Challenges - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Gonzalez Villarreal, O. (2022) 'From Mexico to Milton Keynes', TAS-S Blog, 22 February. Available at: [From Mexico to Milton Keynes - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Lopes, A. (2022) 'From Law to Computer Science', TAS-S Blog, 3 October. Available at: [From Law to Computer Science. - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Moffat, L. (2021) 'ELSI, Workshops, and the things in-between', TAS-S Blog, 2 September. Available at: [ELSI, Workshops, and the things in-between - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Moffat, L. (2021) 'ELSI, Workshops', TAS-S Blog, 16 July. Available at: [ELSI Workshops - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Moffat, L. (2022) 'National Highways Workshop #1', TAS-S Blog, 17 February. Available at: [National Highways Workshop #1 - UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

8. Publications, Media & Products

Blogs

Moffat, L. (2023) 'Backcasting Workshop with National Highways', TAS-S Blog, 26 July. Available at: [Backcasting Workshop with National Highways – UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Wei, Z. (2021) 'PhD to PDRA, TAS-S Blog, 5 November. Available at: [PhD to PDRA – UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Yu, Z. (2021) 'A new job at Lancaster University!', TAS-S Blog, 7 July. Available at: [A new job at Lancaster University! – UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Yu, Z. and Yuksek, B. (2022) 'AI-aided Safety in Urban Airspace: From a Dream to the Real World', TAS-S Blog, 13 October. Available at: [AI-aided Safety in Urban Airspace: From a Dream to the Real World – UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Yukse, B. (2021) 'Autonomous Systems, aviation and TAS-S', TAS-S Blog, 28 June. Available at: [Autonomous Systems, aviation and TAS-S – UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

Yukse, B. (2023) 'TAS-S ESG', TAS-S Blog, 19 June. Available at: [TAS-S ESG – UKRI TAS-S: Trustworthy Autonomous Systems Node in Security \(lancs.ac.uk\)](#).

9. Engagement Activities

Talks

- Talks for TAS-S node, University of Edinburgh, the TAS Governance Node
- Invited talk at Budapest University of Technology and Economics
- Engagement with AI for Collective Intelligence (AI4CI) in Bristol ‘Resilient federated learning: Where performance meets constraints’, the TAS Governance Node
- TAS Showcase 2024 panel discussions around issues of Security and Defence
- TAS-S presentation on Continental and CISPA at VW, Germany
- TAS-S presentation to Cyber Resilience policy team and The National Cyber Security Centre
- TAS-S presentation at NIST/NHSTA Cybersecurity Framework, Washington
- Conference Paper - GMHC-T2M Annual Conference
- Use and Abuse of AI in Autonomous Systems
- Talk on the TAS-S approach to Autonomous Systems Security
- US-Asia-Pacific Space Policy for Autonomous Systems
- Autonomous Systems Security Models
- Human Factors in Road Safety
- Talk to LIRA
- What is Trust anyway?
- The Institute of Engineering and Technology
- New Zealand Institute of Driver Educators National Conference
- ADINJC & Intelligent Instructor National Conference and Expo
- Towards More Interpretable & Greener Deep Learning
- Relational Critiques of Autonomous Systems: Speculating Automated Urban Futures
- TAS-S and National Highways: Creative Methods for Collaboration (seminar for TAS-G node)
- Presentation to TAS Strategic Advisory Network

9. Engagement Activities

Formal Working Group

- STSMN (Science and Technology Studies in the Midlands and North) meeting hosted by the University of York
- TAS-S External Stakeholders' Group Meeting 2023
- Discussion meeting at Centre for Information Security, Hannover Germany
- Autonomous Systems Security Models
- TAS Programme Workshop with Boeing Defence
- TAS Thought Pieces with Thales
- TAS-S Overview seminar (6th April) with Computing-Enabled Networked Physical Systems (CPNS) Interagency Working Group at the US National Coordination Office for Networking and Information Technology (NCO/NITRD)
- European Commission Cybersecurity Roadmap Task group-plenary
- HFM European Technology Summit 2022
- Discussions with Lancashire Police.
- Policy/Standardization Board meetings: ISO 26262
- Meetings with senior management in industry
- Discussions with NHTSA, US DOD and NSF
- GM/BMW/TUV-Germany
- Panel/workshop with Continental/Airbus
- Collaborative discussions with TAS-S External Stakeholder Group Members
- European Commission Cybersecurity Roadmap Task group
- Social Science and the Ethics of Digital Technologies

9. Engagement Activities

Workshops

- Workshop of the TAS NODES community to exchange knowledge and experiences on Equality Diversity and Inclusion (EDI) and Responsible Research and Innovation (RRI)
- The Workshop on Trustworthy and Useful Tools for Mobile Phone Extraction (UKRI Trustworthy Autonomous Systems Hub)
- National Highways Backcasting Workshop
- TAS thought pieces
- TAS Workshop
- TAS '23: The First International Symposium on Trustworthy Autonomous Systems
- Cybersecurity in Space
- Cyber Security Leadership Symposium 2023
- Security Lancaster Seminar
- EC meeting on Secure AS: Technologies and Policies
- Exploring the Space-Cyber Security Frontiers
- AI in Education & ethics through design workshops.
- Talk at Safe and Trustworthy AI Workshop
- Engagement with BAE Systems
- Papers Chair for First International Symposium on Trustworthy Autonomous Systems (TAS '23)
- IEEE ACSOS Regional Event UK
- National Highways 2nd Workshop
- Group seminar
- Group seminar
- External Stakeholders' Group (ESG) Meeting 2022.
- Surveillance Workshop: Use and Abuse of AS for Surveillance
- Workshop on risks and security assurance cases for use of drones in commercial airspace
- NHS Lancashire Healthier Living Programme
- TAS Thought Piece Workshop
- Verification of Autonomous Systems
- Verification of Autonomous Systems
- Delivering the AI Strategy – the use of new AI technologies in industry and the public sector
- TAS-S External Stakeholders' Group Meeting (2021)
- National Highways 1st Workshop
- TAS-S External Stakeholders' Group Meeting 2023

10. Acknowledgements and contact details

This work is supported by the Engineering and Physical Sciences Research Council
[grant number: EP/V026763/1].

We would like to thank our stakeholders for their continued support.



Contact TAS-S Node

TAS-S Node Website:
<https://tas-security.lancs.ac.uk/>

Lancaster University:

Corinne May-Chahal
c.may-chahal@lancaster.ac.uk

Cranfield University:

Weisi Guo
weisi.guo@cranfield.ac.uk