EPSRC Trustworthy Autonomous Systems

The Security Node (TAS-S)

1st External Stakeholders Workshop

March 29th, 2021

https://tas-security.lancs.ac.uk/

https://security.tas.ac.uk

Engineering and Physical Sciences Research Council

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

Agenda

0930	Workshop Introduction (Neeraj Suri, Lancaster University)
	TAS-S Overview
0945-1200	Session 1: TAS-S Research Strands – Activity Plans
0945-1015	**RS1: Securing the AS "Usage" Environment (Lead: N. Suri, Lancaster)** Theme A: Dynamic and Compositional AS Security (N. Suri, Lancaster) Theme B: Explainable & Verifiable Decision Making (P. Angelov, Lancaster)
1015-1100 ▶ 11am Break	**RS2: Securing the AS "Operations" Environment (Lead: W. Guo, Cranfield)** Theme A: Security in the Mission and Operational Surface (P. Angelov, Lancaster) Theme B: Securing the Control Surface (G. Inalhan, Cranfield) Theme C: Securing the Cross-Layer Networking Surface (W. Guo, Cranfield)
1115-1200	**RS3: Securing the AS "Users" Environment (Lead: C. May-Chahal, Lancaster)** Theme A: Behavior Adaptation as a Basis of Security by Design (L. Dorn, Cranfield) Theme B: Organizational Socio-Technical Mitigation (J. Deville, Lancaster) Theme C: Ethics and governance of AS security (C. Easton, Lancaster)
▶ 1200-1245	SIESTA
1245-1400	Session 2: Stakeholder Presentations
1400-1445 ▶ 1445-Break	Session 3: Workshop on Ethical, Legal and Social Issues (ELSI)
1500-1545	Session 4: PANEL "Priorities for AS Security – The Road Ahead"
1545-1630	Closing Session: Observations by Advisory Group and Action Items

The Team

Lancaster University	
Prof. N. Suri (PI)	Systems Security
Prof. P. Angelov	ML/Intelligent Systems
Prof. D. Hutchison	Network Security
Dr. V. Giotsas	Network Security
Prof. C. May-Chahal	Social Sciences
Dr. J. Deville	Sociology
Dr. C. Easton	Law
Pam Forster	Project Manager
Cranfield University	
Prof. W. Guo	Machine Intelligence
Prof. G. Inalhan	AI/Autonomous Systems
Prof. A. Tsourdos	Autonomous Systems
Dr. L Dorn	Behavior Sciences












The most important folks: Our RA’s & PhD students!!!

The Team

Advisory Council: The Base Stakeholders	Academia, Industry, Policy

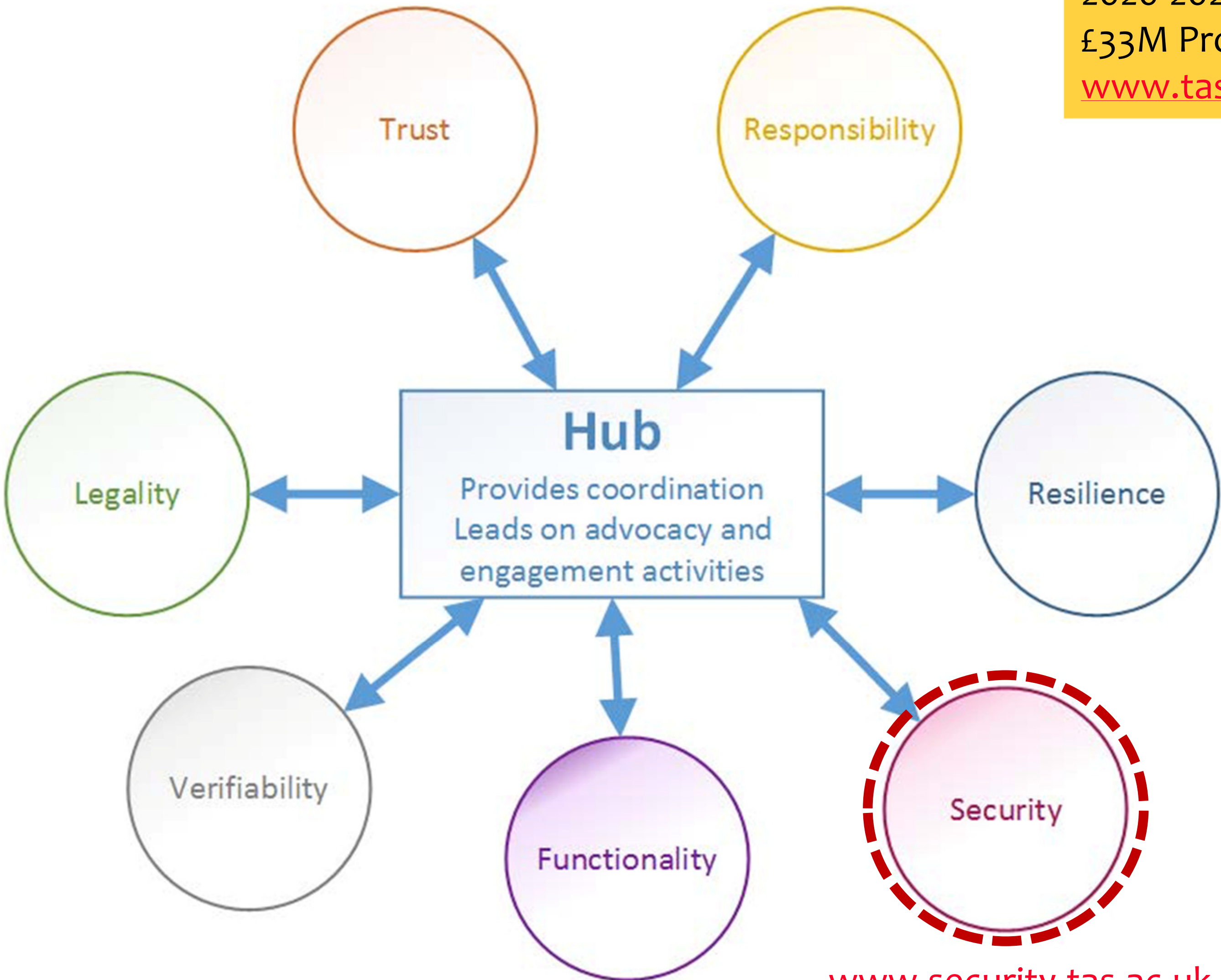
New Stakeholders

Advisory Group: Project	
Prof. Carl Landwehr	CDT/U. Michigan
Prof. Robin Bloomfield	Adelard/City Univ
Dr. Hector Figueiredo	Qinetiq
Dr. Carl Segueira	FlareBright

EPSRC | Hub Liaison	
Dr Victoria Mico Egea	UKRI EPSRC AI/Robotics
Dr. Danielle Lloyd	UKRI EPSRC AI/Robotics
Prof. Gopal Ramchurn	U. Southampton/HUB PI
Prof. Luca Vigano	Kings College London
Prof. Derek McAuley	U. Nottingham
Prof. Jose Such	Kings College London

The EPSRC TAS Program : Strategic Priorities Fund

2020-2024
£33M Program
www.tas.ac.uk



www.security.tas.ac.uk *EP/V026763/1*

Workshop Objectives → Inform + Engage

1. Our TAS-S ideology and research objectives
2. Your opinions, experiences, needs/challenges
3. Discussion across AS "researchers and practitioners"
   - Feedback: Sanity checks, things missed?
   - What types of AS and security risks do you worry about?
   - What aspects of AS [*specification, V&V, perception, control, coordination, communication, use-of-AI,…*] constitute your priorities?
   - Collaboration potential (*use-cases, data, testbeds, validation…*)

Trustworthy Autonomous Systems - *Security Node (TAS-S)*

An Overview

Neeraj Suri

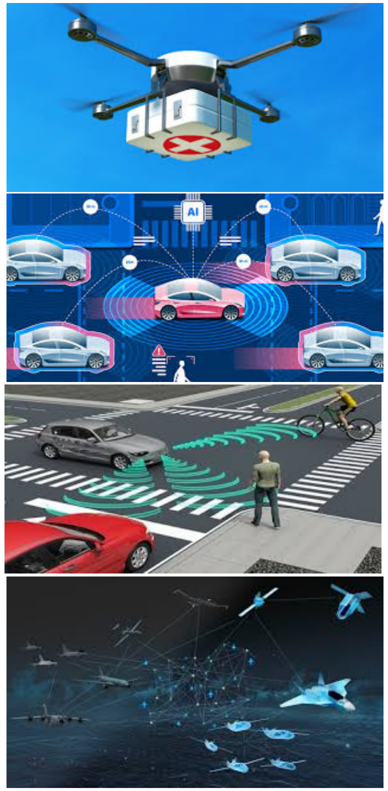
Lancaster University

https://ssg.lancs.ac.uk/people/suri/

https://tas-security.lancs.ac.uk/

Autonomous Systems (AS): Functionality + Scope

Technology to effectively conduct a mission with varied levels of "absence of human intervention" e.g., L0-L5



CPS

- Sensors
- Perception
- Communication
- Control
- Coordination
- Navigation
- Decision
- Adaptation

OODA Loop

Observe
Orient
Decide
Act

Perception, Cognition, Decision

▶ Increasing complexity of applications & environments

▶ Complex connectivity, Complex data streams…

▶ Cognitively/Computationally complex OODA → AI

Trustworthy/Trusted/Trust-in… Autonomy

- Complexity: Things will break, perturbations will happen

  (At all levels of the CPS: AS assets, AS operations, AS environment)

  - Design, mis-configuration, mis-specification, operational: Dependability
  - Bad actors, deliberate disruptive intent: Security

- Technology is (mostly) useful if we can "justifiably" trust it to deliver the "requisite" services
  - Requisite/Correctness is highly subjective
    - Application/context based
    - Tradeoffs across mission, societal, regulatory or economic perspectives

➢ **Aim**: Ensure that the AS (acceptably) delivers the mission!

TAS → TAS-S

- **Complexity is reality** *Assets, Ops, Environment*
- **Uncertainty is reality** *Assets, Ops, Environment*
- **AI is reality**
  Increasingly complex technology dependence beyond human intervenability

➢ ***NEED*: Predictability over/despite Uncertainty**

ASs depend on technology to "base & improve" upon the essence of human experiences, acceptability & regulations to deliver the OODA functionality. This is a very hard problem.

➢ **AS Disruptions are reality**
Increasingly complex, increasingly inter-connected, increasingly attackable

For ASs to provide for "safe+secure" *[predictable]* delivery with *degraded or compromised* systems *[increased uncertainty]* is an even harder problem.

Security → Autonomous "System" is not compromised

- AS assets do not get compromised
- AS ops/mission does not get compromised
- AS user/usage environment does not get compromised
  (Societal spaces: users, regulatory, ethical, collateral damage …)

The mission proscribes the level, acceptability and responses to the compromises!

Security: The Abstract View

1. Given a set of assumptions
2. Create a model of reality (assets, mission, env + threats)
3. Assert a requisite security property

✓ Deploy in the real world (and keep fingers crossed ☺)

<u>Assumptions</u>:

- That our assumptions are valid and complete
- That our models are valid and complete
- That the AS + environment + attackers behave as modeled!

<u>Security</u>: Compromise of the Assumptions or the Models

An attacker can use/abuse/ignore/subvert assumptions & models

Security: The Reality in an AS

- Accurate & complete model of system, mission, environment?
- Accurate & complete sensory streams?
- Accurate & complete perception/cognition/decision +AI?
- Accurate & complete specification of the threats across the socio-technical attack surface? (UU)
- Accurate & complete specification of post-attack information streams, resources, decision options?
- Accurate & compete specification of user/usage aspects?

AS: A world of Uncertainties!

AS Attack Surfaces & Dynamic Responsiveness

- Complex attacks – discrete, collusion, multi-layered
- Dynamic & complex - mission and societal – operational environments + corresponding diverse attack surfaces
- "Adaptive & run-time" OODA decisions with incomplete and uncertain data streams and resources
- Dependence on non-deterministic AI technologies

AS: Predictability despite Uncertainty

AS (after attacks): Predictability despite increased Uncertainty!

<u>Security</u>: Works best in structured environments

<u>AS</u>: Dynamic, adaptive + <u>users</u>… anything but structured

How do we provide well-structured AS security in complex socio-technical mission environments that are inherently unstructured?

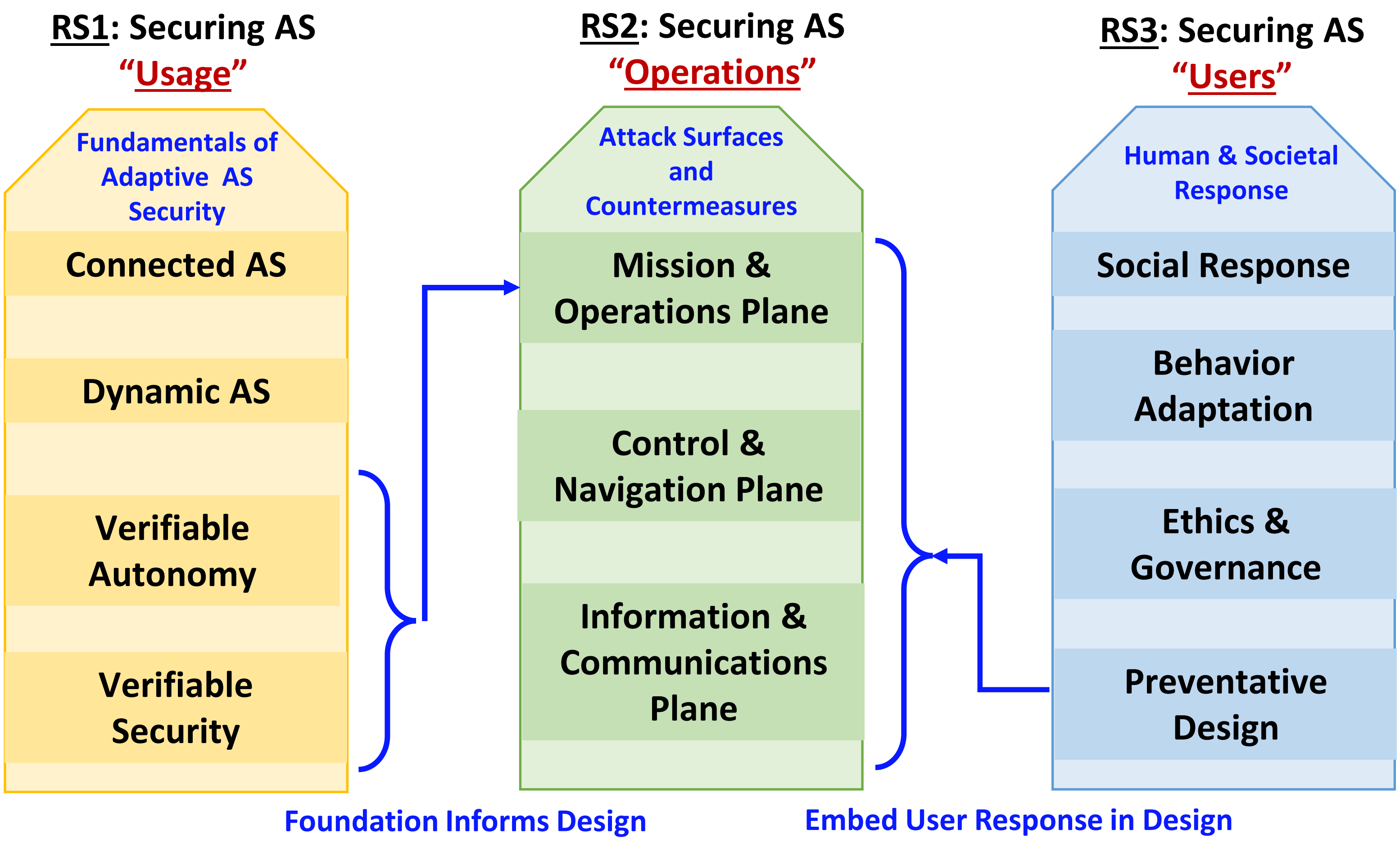
- We know how to provide (partial & expensive) point solutions.

- What we critically lack is a scientific framework that can provide *"composable, scalable & verifiable" mission adaptive* socio-technical security! **"Predictability despite Uncertainty"**

Challenge: Unstructured, Uncontrolled, Dynamic Environment

- Can we secure the AS Usage basis?
  - Foundations: Specify, Compose, Explain, Verify
- Can we secure the AS Operations?
  - Ascertain & Mitigate Threats: Mission, Operations, Control, Comm
- Can we secure the AS User spaces?
  - Behavior adaptation, Ethics, Regulatory, Governance

TAS-S Research Strands (RS)



Basic Research → Applied → Testbed Validation

Research Strand (RS1): Securing the AS "Usage"

Fundamentals of Adaptive AS Security

Connected AS

Dynamic AS

Verifiable Autonomy

Verifiable Security

Theme A: Dynamic and Compositional AS Security

(N. Suri, A. Tsourdos, G. Inalhan)

Theme B: Explainable & Verifiable Decision Making

(P. Angelov, N. Suri, W. Guo, G. Inalhan)

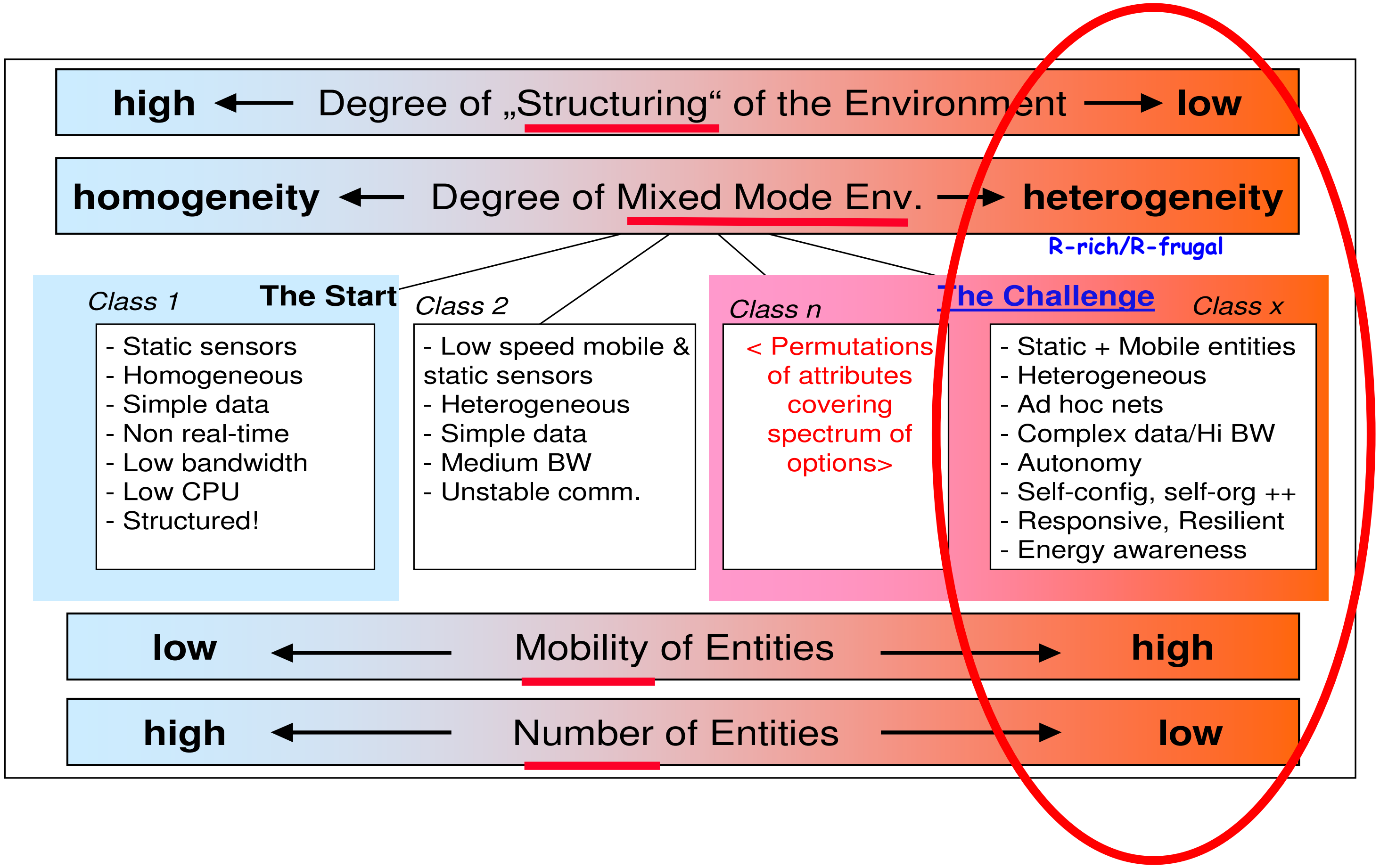
RS1 Scope

RS1 Objectives	State-of-the-Art and Gaps	Innovation Target
Compositional & Verifiable "System of Systems" Security	Security approaches well-developed in structured space, but AS operate in dynamic unstructured environments	Develop dynamic & adaptive AS security measures (specification, composition, verification) responding to multi-modal and uncertain threats

RS1: Foundations of AS Security – The Environment

The needs: Scalable, Composable, Verifiable Security (**Structured**)

The problem: Variety, Volume, Velocity… **(Unstructured)**



RS1 Theme A: Specification and Models

The issues: Specification of "Uncertainty"

- How do we specify the AS systems environment?
- How do we specify the AS security specs?
- How do we compose security? (Complex collaborative – SoS)
- How do we form, adapt and solve AS models… on the fly?
- How do we specify verifiable (offline/online) AS behavior?
    - AI is wonderful, but deterministic reproducibility is not its strength

RS1A Target 1: Establishing Security Specifications

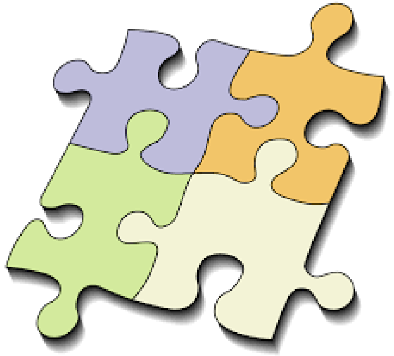
Security: Given a system's "specifications" of assumptions, models and threats, assert a measurable security property

Approach: Bounding uncertainties to ensure predictability

- Ascertain security attributes per operational plane of the AS
- Ascertain dependencies on security attributes/threat models
- Ascertain minimum environment characteristics and the tolerances needed to "sustain" a security attribute

Challenge: The AS environment, ops and threats – all are dynamic!

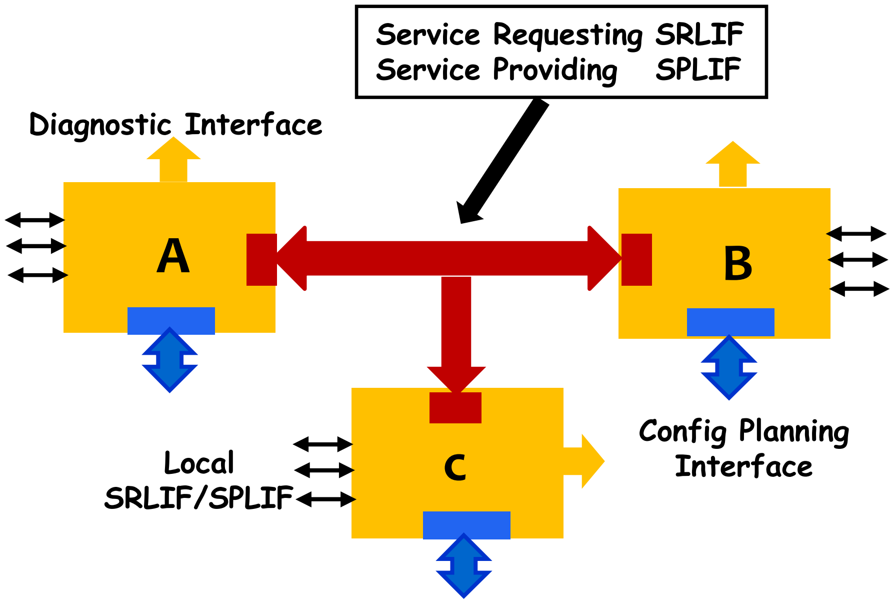
RS1A Target 2: Collaborative AS → SoS Compositions



- The pieces need to fit
  - Functionality needs to compose (invariance & growth): 2 + 2 ≥ 4!
  - Threat models need to compose: No leaks or new threats
  - Security properties (+ metrics) need to compose: C.I.A +++
- Compositions result in "emergent" behaviors
  - "Emergence" in <u>not</u> a popular word in security

Compositions: Linking Interface Model (LIF)

- Specifying AS "components"
- Specification of functional properties: values, timing, resource constraints…
- Specification of non-functional properties: FT, security…
- Specification of security metrics
- Composition rules
- Is a component/interface stateful or stateless?



LIF State & Component States can differ

RS1A Recap

Target: Fundamentals of adaptive AS security specifications to achieve "Predictability despite Uncertainties"

Progressive Outcomes

- Specification framework characterising the relationships across the dynamic and unstructured AS environment & security attributes
- Compositional framework for collaborative, disruptive and scalable security
- ➢ Run-time security policy framework for AS

Open areas (also as a basis for collaboration)

- What AS models + security attributes really matter in reality?
- What problems does the community encounter over collaborative AS?
- Repository of synthetic AS deployment scenarios?

RS1: Fundamentals of Autonomous Systems Security

Theme B:

Verifiable Autonomy and Security

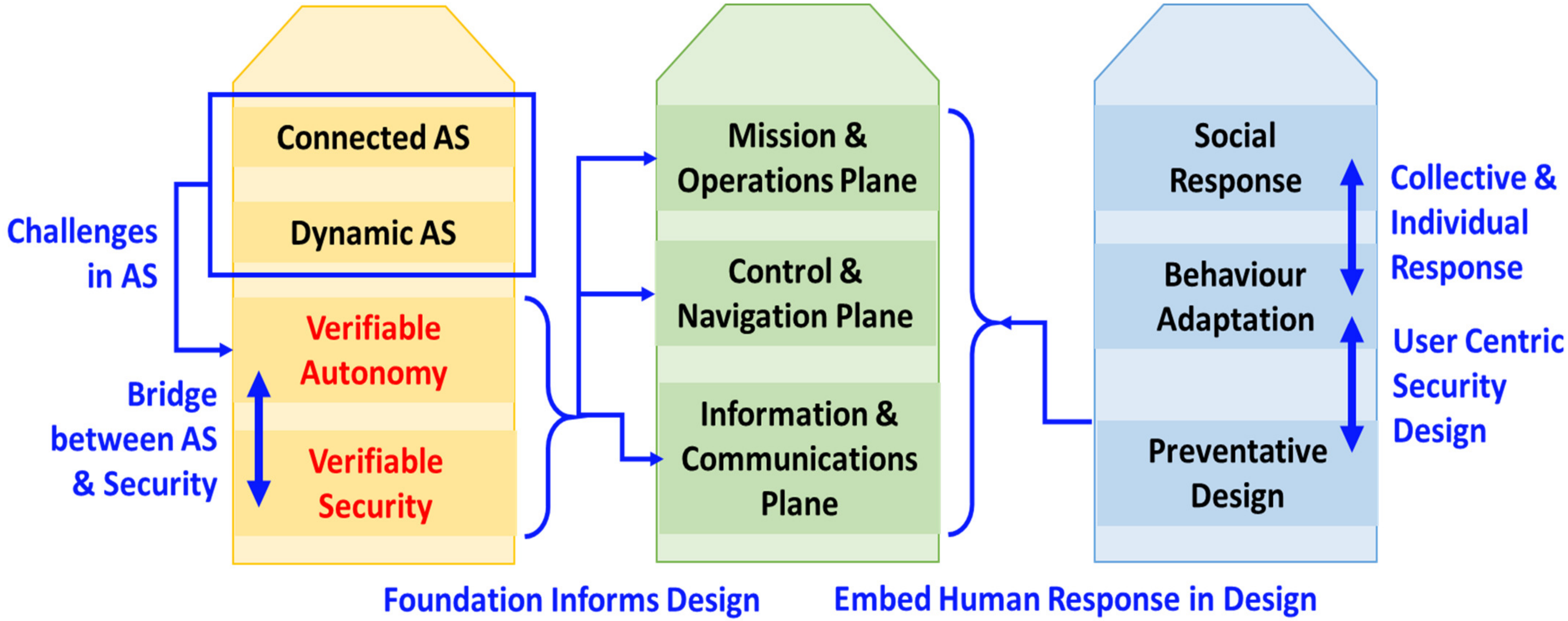
Plamen Angelov (Lancaster Univ)

Neeraj Suri (Lancaster Univ)

Weisi Guo (Cranfield Univ)

Gokhan Inalhan (Cranfield Univ)

Trustworthy Autonomous Systems – Security Node



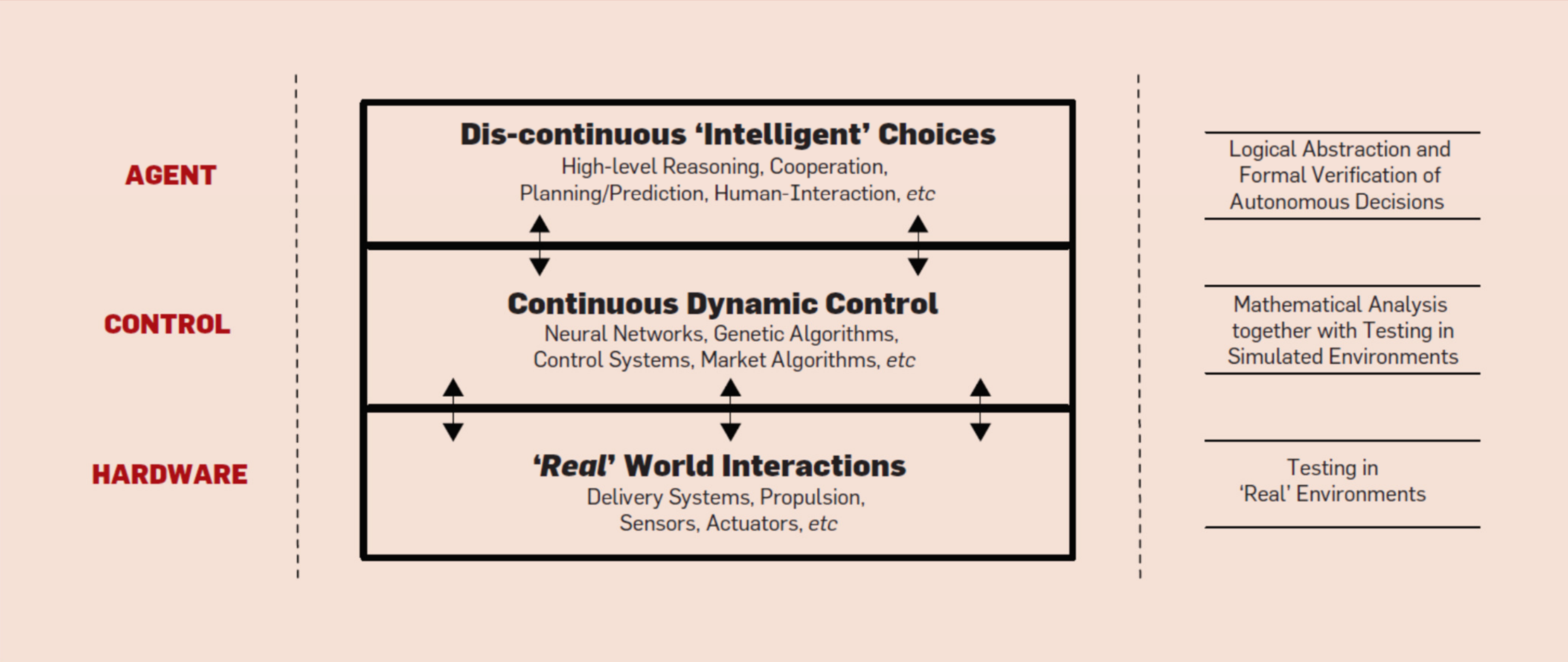
RS1: Fundamentals of AS Security

Autonomy has to be verifiable (deterministic?)

- Assured Autonomy
- Known unknowns
  - Identify vulnerabilities
  - Verifiable countermeasures
  - Formal methods may be applicable
- Unknown unknowns, unexpected
  - Detect, recognize, learn from unexpected
  - Bounded performance, egress routes, mission abort
  - Explainable by design deep learning, exploratory classifiers (xClass)
- Proliferation of AI and ML (often non-deterministic) raises questions related to (deterministic) verification

RS1: Fundamentals of AS Security

Figure 1. Typical hybrid autonomous system architecture—with suitable analysis techniques noted.



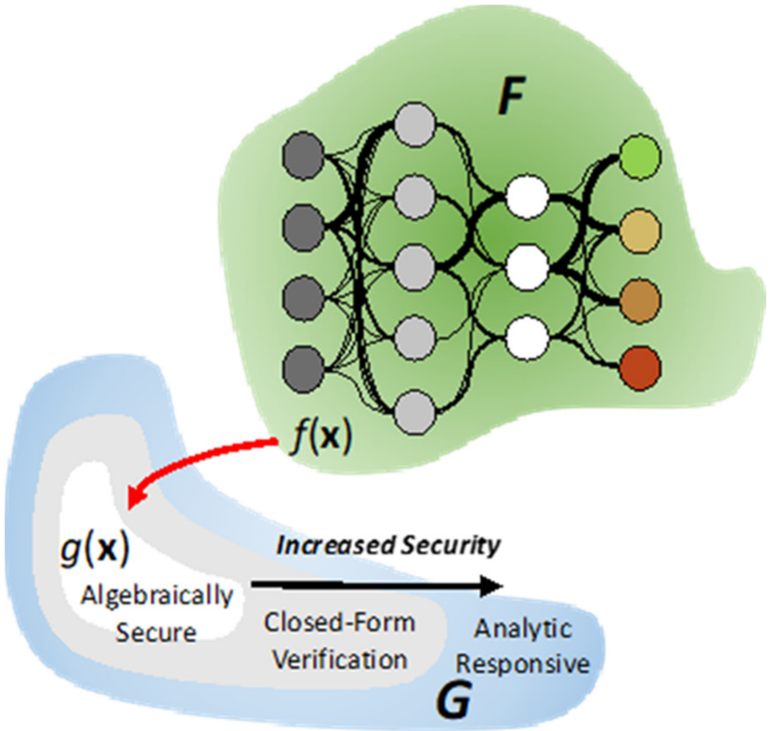
M. Fisher et al., Verifying Autonomous Systems, Communications of the ACM, 56(9): 84-93, Sept 2013

RS1: Fundamentals of AS Security

Autonomous Systems need to be secure:

- Against external threats (environment, adversaries)
- Against internal threats (system itself, e.g., **algorithm**, communication, insider threats)

➢ Interpretable deep learning with verifiable proofs



RS1: Fundamentals of AS Security

Characteristics:

- Context related – mission plane
- System related – control, navigation, machine health
- Network facing – information, data/sensors
- Human related (even though autonomous – part of a system of systems)

Challenges and open questions:

- Open/dynamic operational environments (how to factor/specify/model subject to "unknown unknowns")?

- Difficult to elicit formal requirements for complex missions (completeness? dynamic specifications?)

➢ How about amorphous models such as neural networks, deep learning and, more generally, learning and adaptation algorithms?

Challenges and open questions:

- Heterogeneity of AS
- Design time vs run time verification?
- How about runtime performance under uncertainties?
- Full guarantees of safety or graceful degradation and egress?

RS1: Theme B AI, Deep Learning and Security

AS are increasingly using and relying on AI and various forms of machine learning including deep learning (DL)

This creates opportunities for performance but opens the door for security treats and vulnerabilities.

For example:

- Uncontrolled high dimensional (HD) noise or adversarial data attacks are difficult to expose at HD levels of DNN

- The research is divided into developing both real-time data-driven defences, and statistically grounded certificate defences

Examples:

Adversarial Training, robust stochastic gradient descent (SGD) tackles corrupted data or gradients during the training phase by checking for adversarial examples.

However, this does not effectively deal with real time backdoor access to training data that may add wrong data or labels

- This empirical approach does not offer guarantees, certificates
- Certificate Filters: offer proofs to what attacks can be countered using statistical guarantees integrated into the DNN.
- Traditionally, in low dimensional data, we can identify corruption/noise through covariance checks.

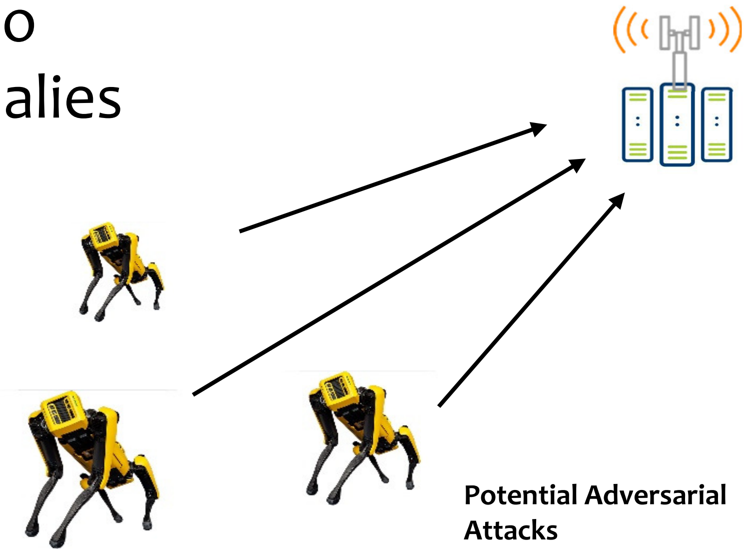
Review of DNN Security & Defence

- This becomes more challenging at HD, especially with mixed data types and mixed adversarial statistics.

- Other certified defences that might not operate in real time include:

  1) randomised smoothing with soft classifiers, and

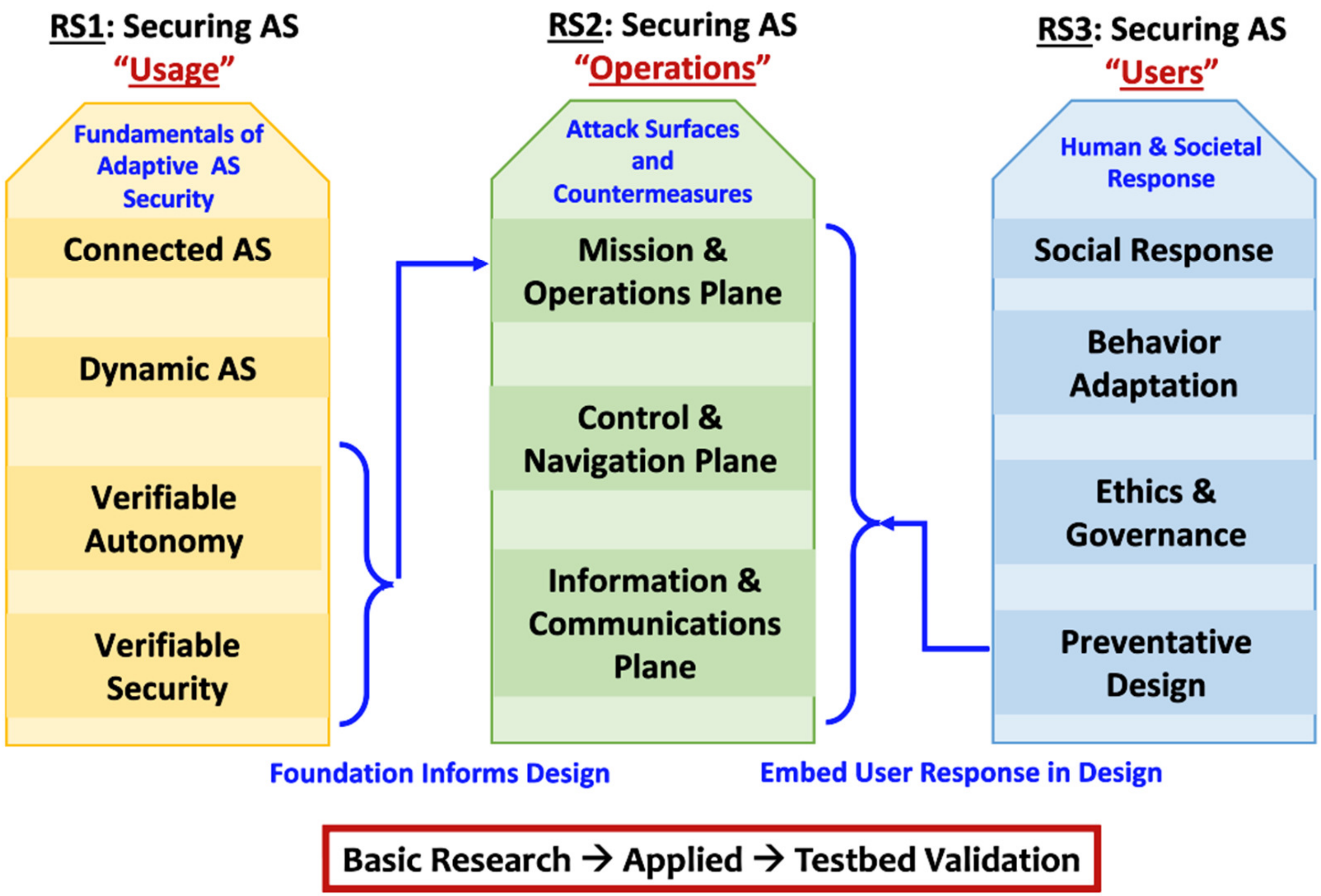
  2) manifold based defences to identify data topology anomalies



RS-2: “Usage Environment”

Prof. Weisi Guo (RS2 lead, Network lead)
Prof. Plamen Angelov (Mission lead)
Prof. Gokhan Inalhan (Control lead)
Prof. Antonios Tsourdos
Dr. Vasileios Giotsas
Prof. David Hutchison

Operation Space

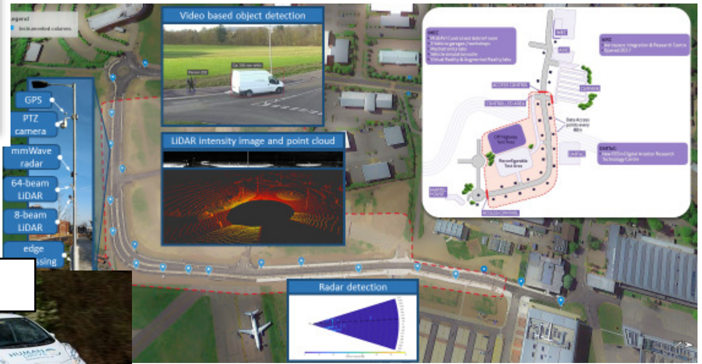


Real Autonomous System Test Capability (Theory to Practice)

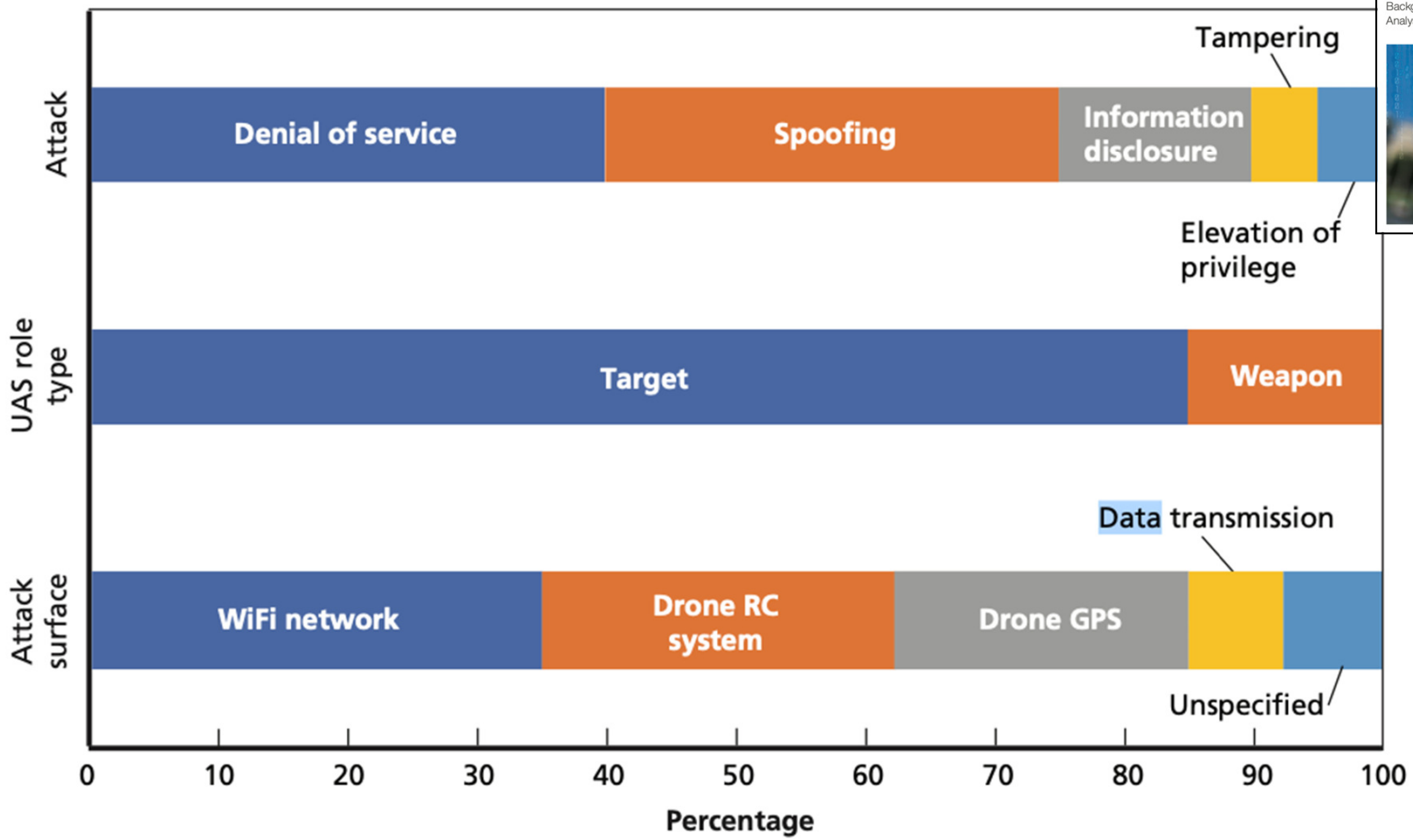
Global Research Airport & Airspace (only 1 in world) with Queen's Award UK flying laboratory

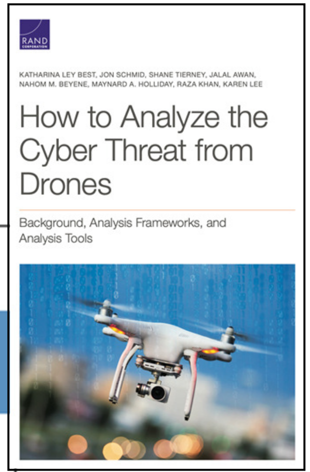
UK National Unmanned BVLOS Drone Corridor

Intelligent Air-Ground Joint Autonomy Testing



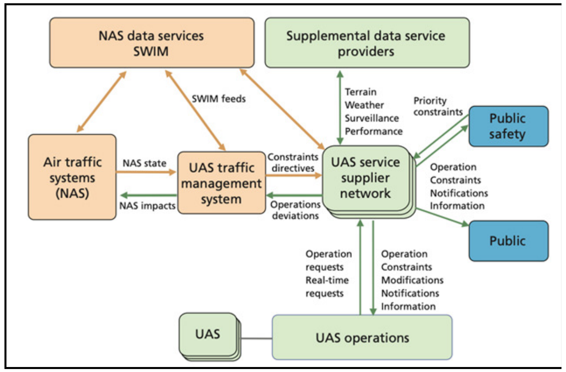
Real Autonomous System Attack Statistics





Real Autonomous System Attack Vectors & Ecosystem

Trend	Key UAS Feature	STRIDE Taxonomy Threat	Vulnerabilities and Attack Vectors
Simplified Control and Operation	Camera view-based flight; following target on camera	Repudiation and Information Disclosure	Third-party monitoring of user activities
	Gesture and speech-directed flight control	Elevation of Privilege and Tampering	Alteration of factory-installed configurations
Self-Operation and Vigilance	Location or sensor-based payload manipulation (e.g., crop spraying, medical supply delivery)	Elevation of Privilege	Intercept of payload usage or delivery
	Swarm drone maneuvers; multi-UAS operations	Elevation of Privilege and Tampering	Scaled-propagation of operational errors
	Preplanned hovering; patrol routines	Spoofing or Tampering	Override of authentic GPS signal or uploaded navigation files
Self-Maintenance and Protection	High-speed obstacle avoidance	Spoofing and Denial of Service	Sensor saturation or interference for obstruction of "view"
	Auto-docking; recharge; return to home	Repudiation and Information Disclosure or Spoofing and denial of service	Third-party monitoring of user activities and sensor interference for failure to register "home" state



State of the Art and Innovation

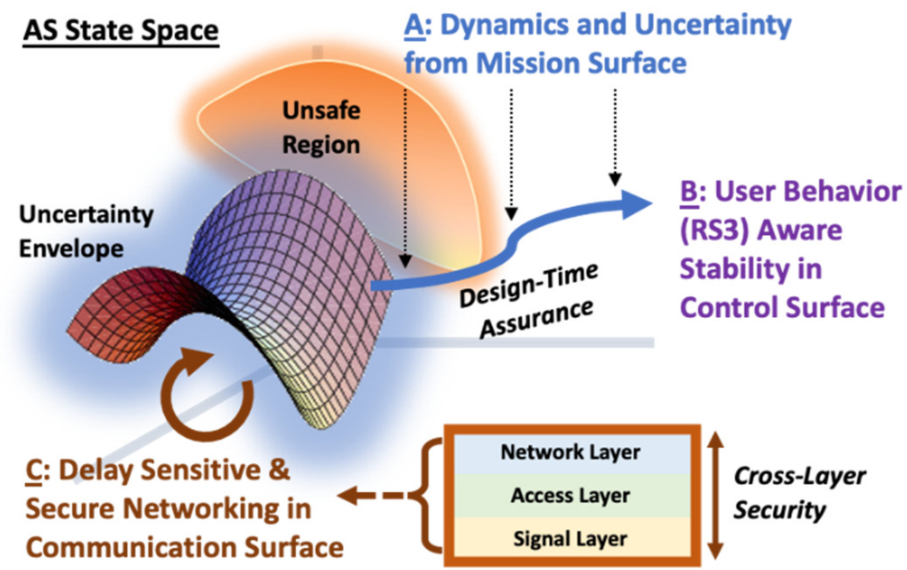
RS Objectives	State-of-the-Art and Gaps	FASMAS Innovation
RS1: Compositional & Verifiable "System of Systems" Security.	Security approaches well developed in structured space, but ASs operate in dynamic unstructured environments.	Develop dynamic/adaptive security measures for AS responding to multi-modal and uncertain threats.
RS2: Multi-layer attack surface mitigation.	Mostly discrete layer analysis. Integrated mitigation of cascaded cross-layer threats in a dynamic AS space is in its infancy.	Hybrid cross-layer mitigation across mission, control, and information layers for AS operations.
RS3: Adaptive Socio-Technical and Legal risk mitigation.	Limited studies of long-term AS behavior adaptation and integrating technical and user-in-the-loop AS security.	Individual, organizational, and legal adaptation to improve socio-technical security (input to RS2).

State of the Art and Innovation

A. Exposure to cyber-physical attacks by characterizing the attack surfaces, i.e., entry points and likelihoods across the mission surface in a technology & mission-invariant manner.

B. Provide quantifiable safety and feedback to the mission surface when the limits of secure controllability are compromised within a time horizon under current policies and adversarial situations.

C. Provide secure communications across the different layers in the informatics plane from detection of signals to networking.



RS2: Attack Surfaces and Countermeasures

Securing the AS Operations Environment

Theme A:

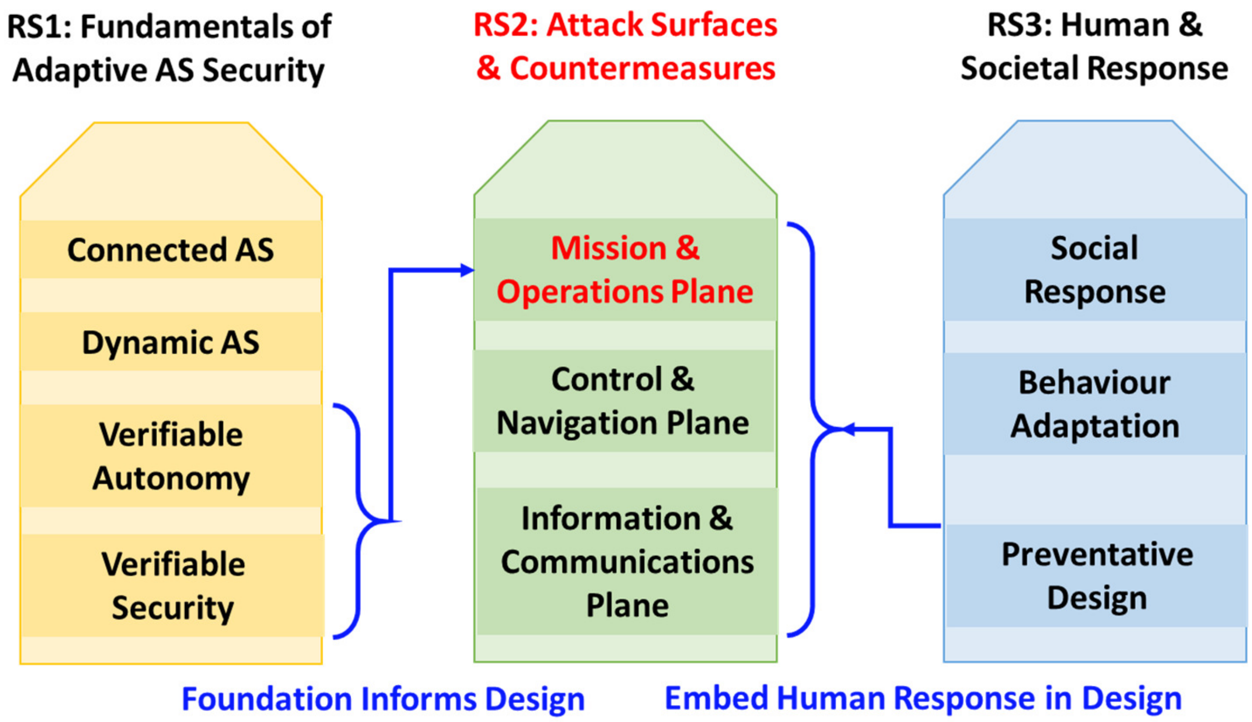
Mission and Operations Surface

Prof. Plamen Angelov (Lancaster Univ)

Prof. Antonios Tsourdos (Cranfield Univ)

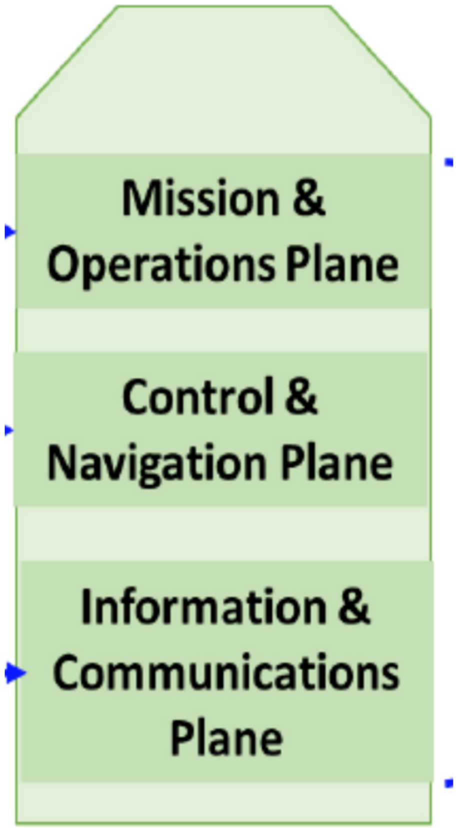
RS2 Theme A: Mission and Operations Surface

Linkages across Research Strands



RS2 Theme A: Mission and Operations Surface

Attack Surfaces and Countermeasures



Mission Surface:

- High level strategic goals, plans, memory/data sets, knowledge, taxonomy and ontologies well as world models get attacked and compromised
- Focus: Mission vulnerabilities, threats & attacks

Operations Surface:

- Tactical - dynamic ops & environment aspects → concerns decision algorithms & mechanisms
- Focus: Operational vulnerabilities, threats & attacks

RS2 Theme A: Mission and Operations Surface - Aims

Ascertain *exposure to* ***cyber-physical attacks*** *by characterizing the* ***attack surfaces****, i.e.,* ***entry points*** *and* ***likelihoods*** *across the mission surface in a technology & mission-invariant manner*

- Identification of attack surfaces and development of mitigation strategies
- Develop algorithms to detect and mitigate threats across the relevant surfaces of AS
- Monitor and guard the mission
- Functional decomposition of AS operation planes
- Complexities related to swarms and network-centric scenarios

RS2 Theme A: The Mission Plane

Dynamics and uncertainty related to the Mission Plane

- Characterizes the essence of an AS to autonomously execute a mission, including element of **coordination** (across the AS entities and/or with the environment)

- The **decision planning** operations that accomplish the mission and the **sensory data streams** supporting navigation, orientation, pattern recognition, including vision, ISTAR, situation awareness, self-organisation, egress conditions for safe/secure fallback

RS2 Theme A: The Mission Plane

Dynamics and uncertainty related to the Mission Plane

- Likely security vulnerabilities include:
  - multi-source sensory data and computations;
  - distribution of the system elements, on-the-ground versus on-board AS task performance

- Furthermore, **inherent uncertainty** in the decision plane contributes to additional security vulnerabilities and may jeopardize mission success

RS2 Theme A: The Operations Plane

Dynamics and uncertainty related to the Operations Plane

- Covers the realizations of the AS protocol, decision and coordination functionality where most AS security compromises (on access control, confidentiality, integrity, availability) transpire
- The new AS challenges are mobility, heterogeneity and **dynamic aggregation** across AS entities
- The approach of identifying the attacks surfaces for AS coordination protocols and execution middleware will be based on the exposition of the knowledge base on distributed systems security approaches and federated learning

RS2 Theme A: AI, Deep Learning and Security

AS are increasingly using and relying on AI and various forms of machine learning including deep learning

This creates opportunities for performance, but opens the door for security treats and vulnerabilities; for example:

in addition to the methods mentioned in RS1B, also

explainable by design deep learning

RS2 Theme A: AI, Deep Learning and Security

- Explainable-by-design forms of Deep Learning offer
  - not only more human-understandable internal working of complex and efficient algorithms of high performance,
  - but also added level of security because the move away from the "black box" nature the mainstream deep learning offers
- It can be used for classification algorithms, for decision making as well as for exploration of a new environment
- Security threats and countermeasures will be studied and analysed both in design and run time

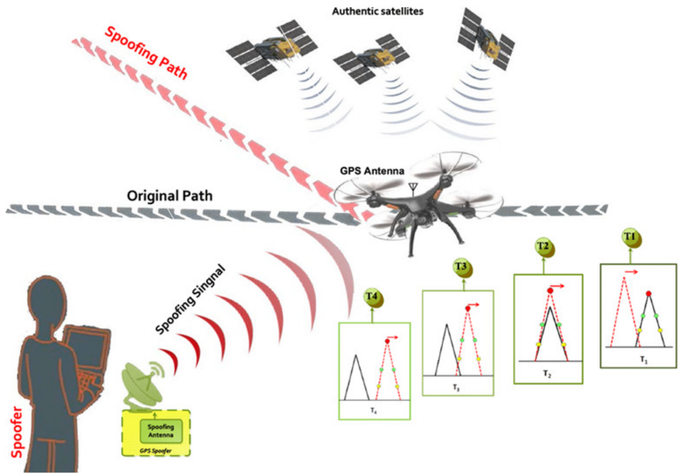
RS-2 B: "Securing the Control Surface"

Prof. Gokhan Inalhan (Control lead)
Prof. Plamen Angelov
Prof. Antonios Tsourdos

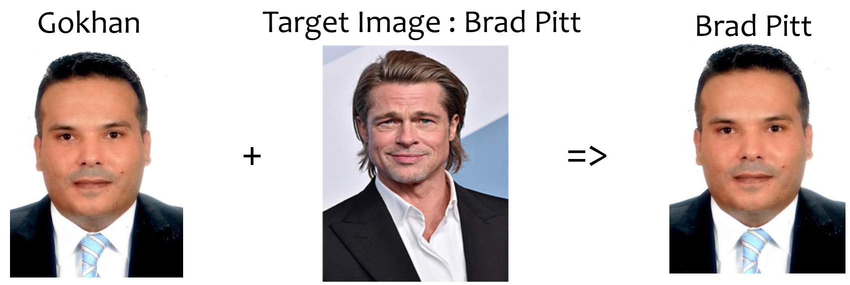
- Autonomous Systems rely on the ability to conduct **run time adaptations of control decisions** over attacks or “perceived” attacks:
  - Adversaries
    - Physical
    - Information-plane
  - Information and dynamic environment uncertainties
  - Degraded performance
    - CNS and Infrastructure
    - Actuators

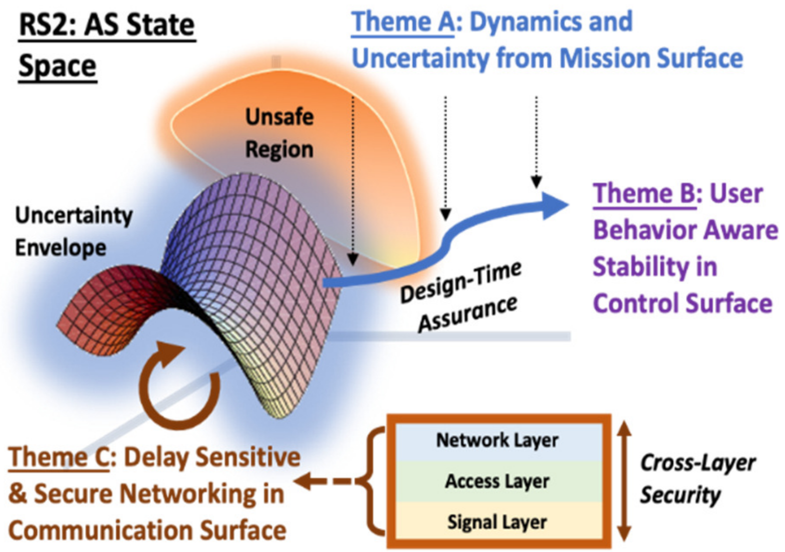
- How to do this in a “**trustworthy**” fashion?
  - Safe
  - Secure
  - Reliable

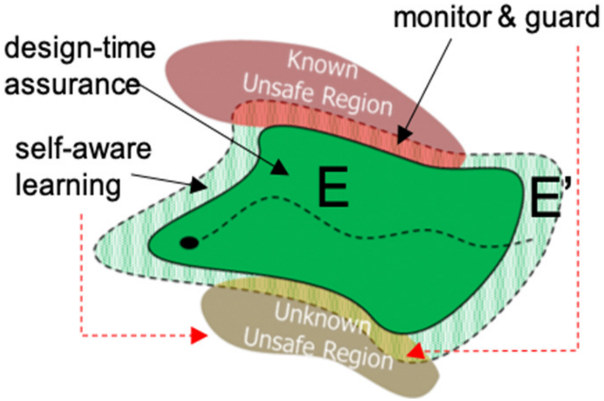
- Sensing and COMM errors
- Loss of an actuator
- Environmental conditions
  - Wind
- Electronic Attacks
  - Jamming
  - Spoofing
- Electromagnetic deception
  - false/duplicate target generation



- Generative Adversarial Networks
  - DNN perception and classification
- Injecting false patterns into data

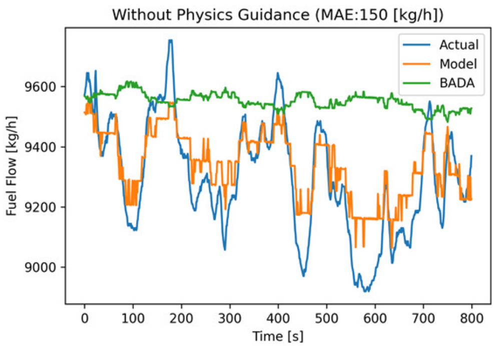




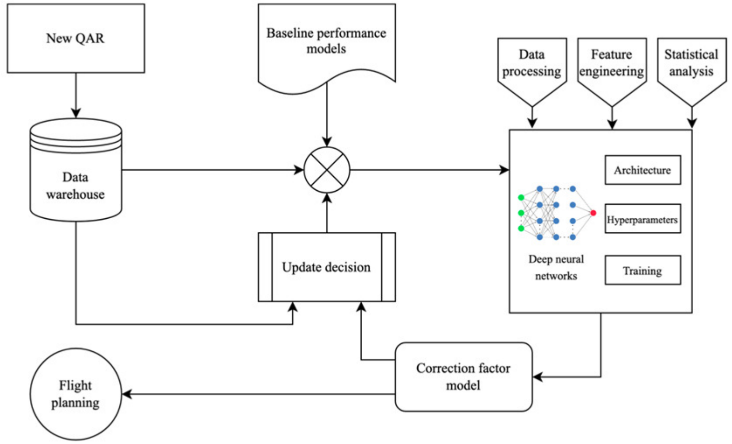


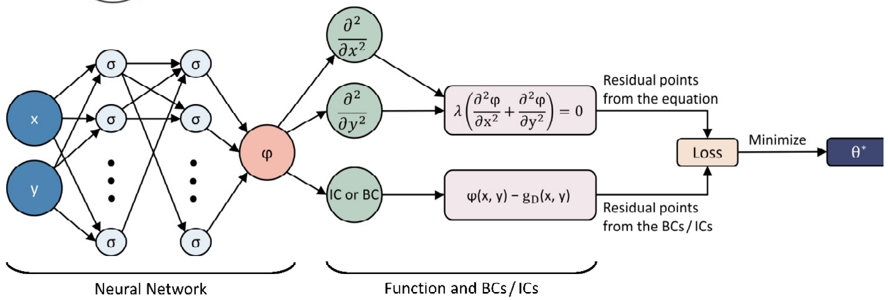
- Provide **quantifiable safety and feedback** to the mission surface when the limits of secure controllability are compromised within a time horizon under current policies and adversarial situations.
- Key Solution Cornerstones in Learning-Enabled Context
  - **Interpretability ➔** Explainable and Trustworthy AI
  - **Continual Assurance ➔** Dynamic Verification & Validation
  - **Adaptive Security Strategies**

- Leading to **Explainable AI**
- Physics Informed Deep Learning
  - Ability to identify system behavior
  - Generalization capability beyond training data input and output sets
  - Ability to detect/classify information and anomalies
    - Degraded performance

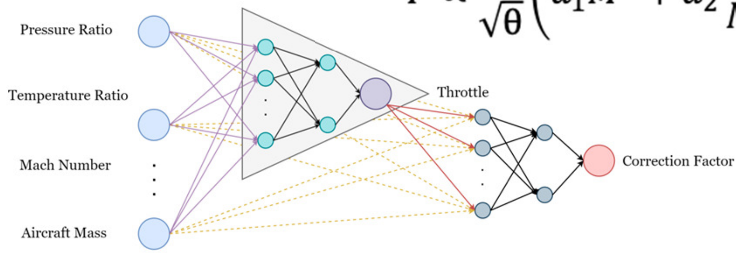


Uzun M, Demirezen MU, Inalhan G. Physics Guided Deep Learning for Data-Driven Aircraft Fuel Consumption Modeling. Aerospace. 2021; 8(2):44.

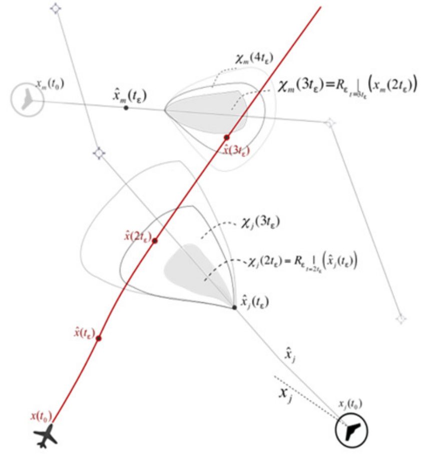


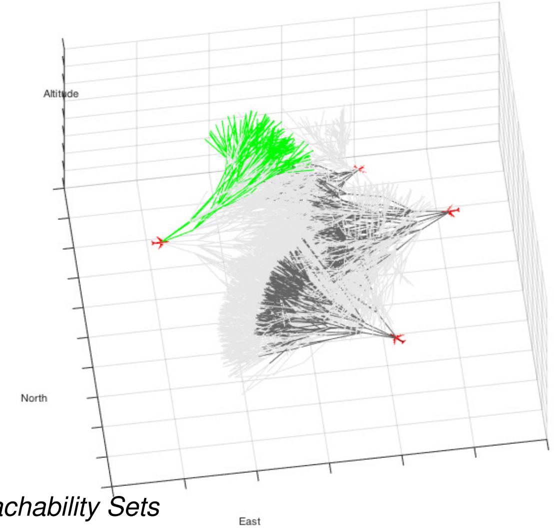


$$F \propto \frac{M}{\sqrt{\theta}}\left(a_1 M^2 + a_2 \frac{m^2}{M^2\delta^2}\right)$$

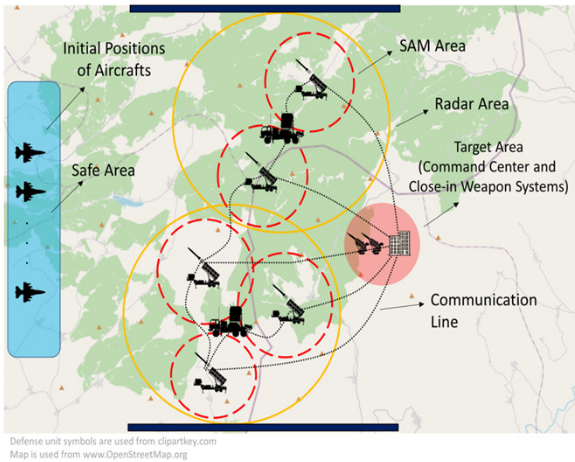


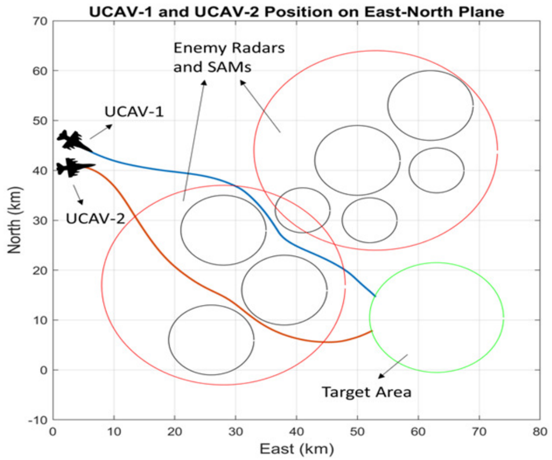
- Dynamic Reachability Sets
  - Detect and Avoid
  - Learning Enabled Context

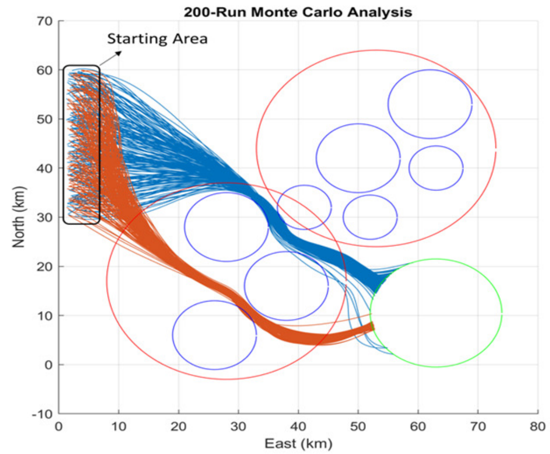




Dynamic Reachability Sets with information uncertainty



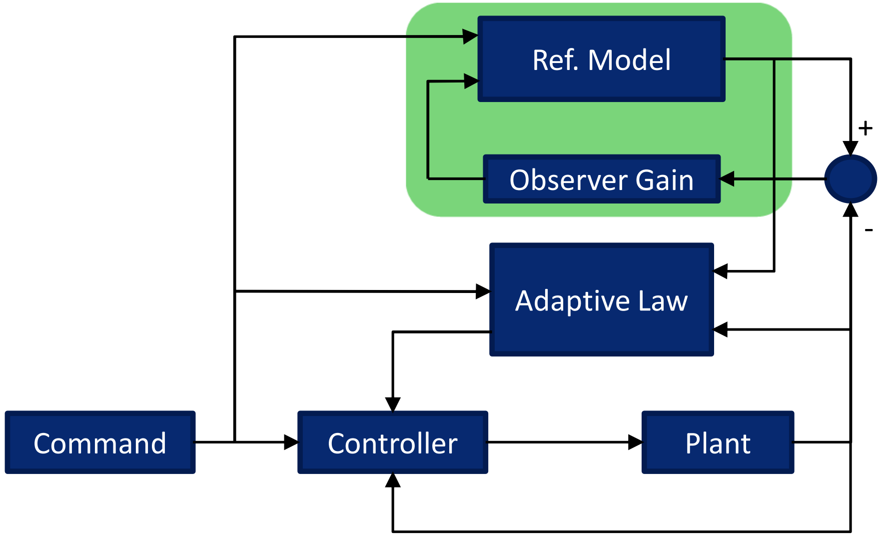


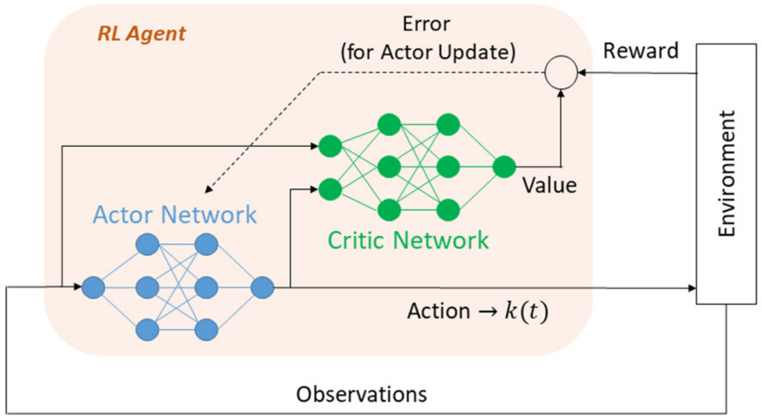


Yuksek B., Demirezen U., Inalhan G., Tsourdos A., “*Centralized Cooperative Path Planning for an Unmanned Combat Aerial Vehicle Fleet Using Reinforcement Learning*”, AIAA Journal of Aerospace Information Systems, 2021, in review

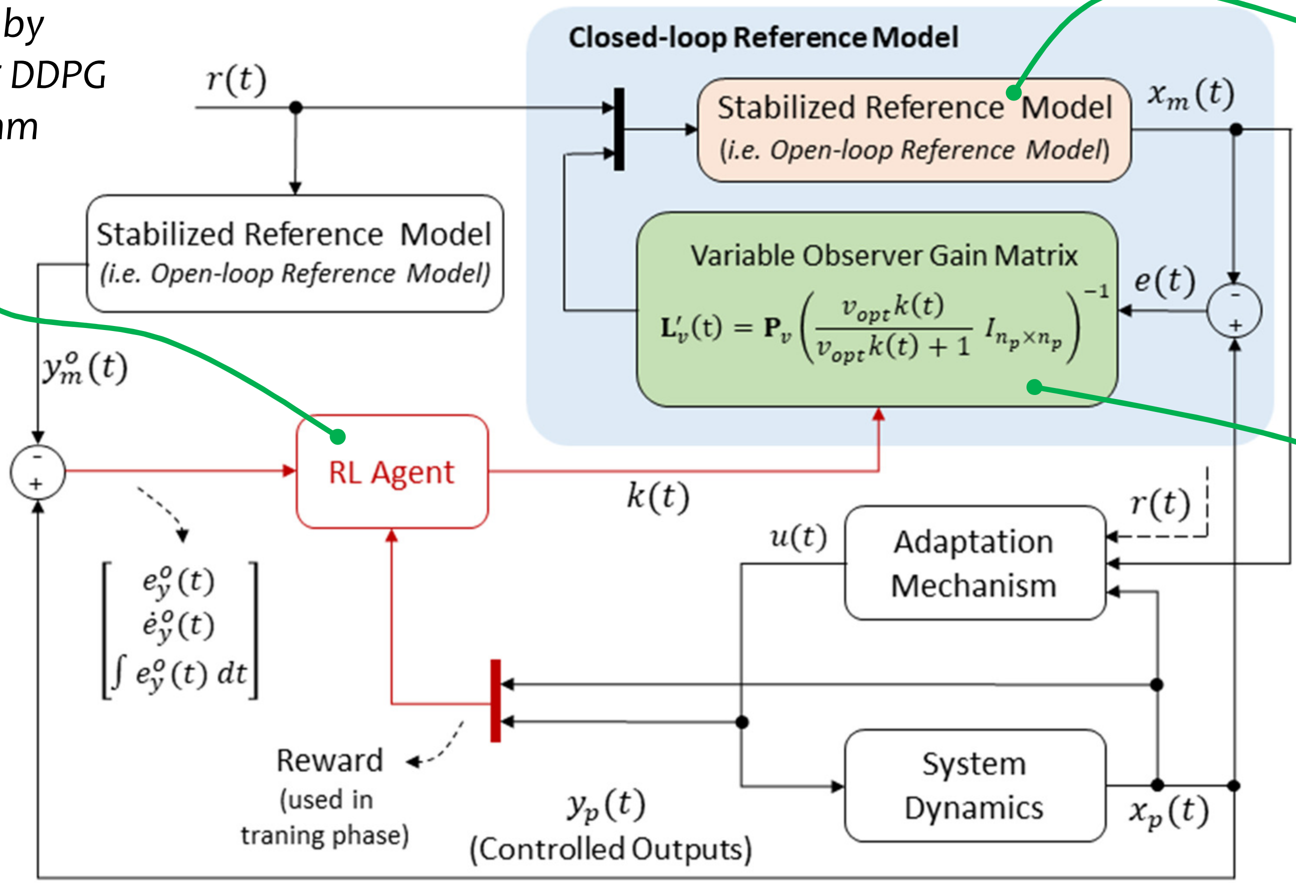
- Deep Reinforcement Learning Based Adaptive Controls
  - Learn adaptation strategy through observation between reference model and the reality

Yuksek B, Inalhan G. Reinforcement Learning Based Closed-loop Reference Model Adaptive Flight Control System Design. International Journal of Adaptive Control and Signal Processing. 2020;1–21.



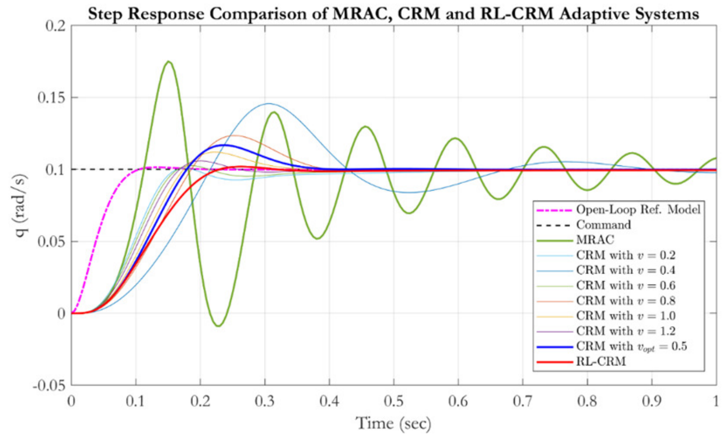


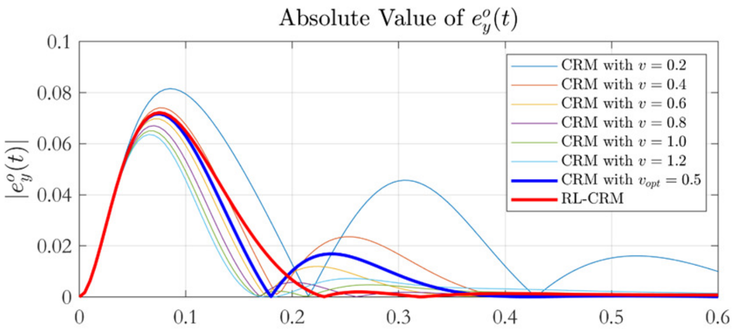
Actor-Critic Structure Trained by utilizing DDPG Algorithm

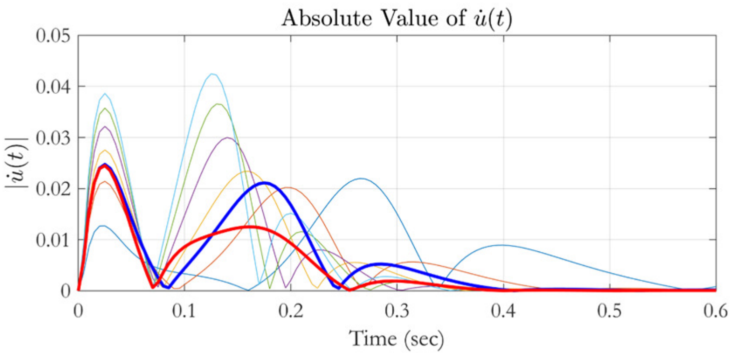


Stabilised model is required if the open-loop dynamics is unstable

Time-varying $k(t)$ provides scaling policy of the observer gain parameter v_{opt}



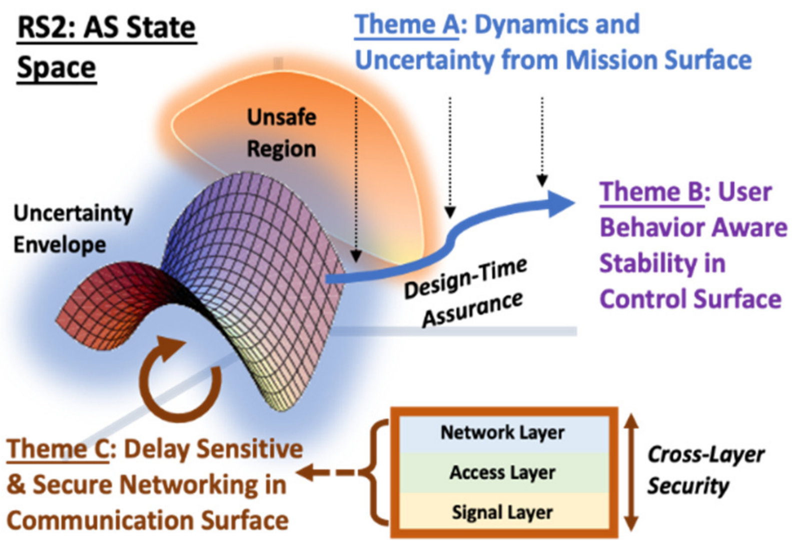




Performance Metrics	MRAC	CRM	Improvement (%)	RL-CRM	Improvement (%)
$\overline{\lVert \dot{\hat{K}}_x \rVert}$	15.2114	3.7341	75.4520	2.4489	83.9008
$\overline{\lVert \dot{\hat{K}}_r \rVert}$	18.4647	7.8298	57.5958	5.5146	70.1344
$\overline{\lVert \dot{\hat{\theta}} \rVert}$	0.0888	0.0338	61.9369	0.0207	76.6892
$\overline{\lVert y_m \rVert_\infty}$	0.2	0.2064	-3.2	0.2	-
$\overline{\lVert e_y \rVert}$	0.4616	0.1957	57.6039	0.1379	70.1256
$\overline{\lVert e_y^o \rVert}$	0.4616	0.3928	14.9047	0.3886	15.8145
$\overline{\lVert \dot{u} \rVert}$	6.5704	2.0811	68.3262	1.4163	78.4290

- **Interpretability ➔** Explainable and Trustworthy AI
- **Continual Assurance ➔** Dynamic Verification & Validation
- **Adaptive Security Strategies**

- **How can we engage with you?**
  - Specific problems
  - Use cases/applications
  - Data
- **What is your expectation from us?**
- **Mechanisms to engage with you?**



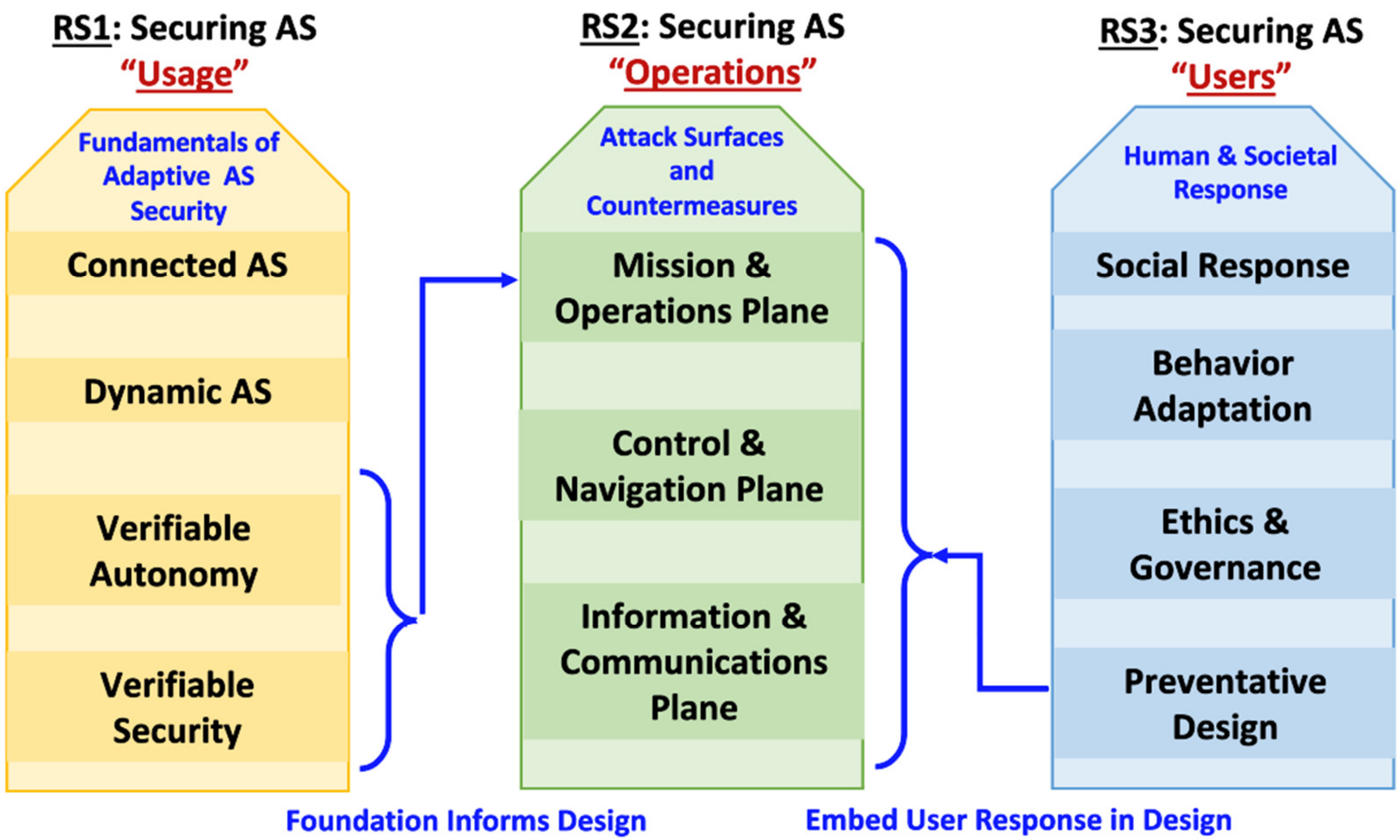
RS-2 C: "Securing the Cross-Layer Networking Surface"

Prof. Weisi Guo (lead) [Physical Signal Security]

Dr. Vasileios Giotsas [Network Security]

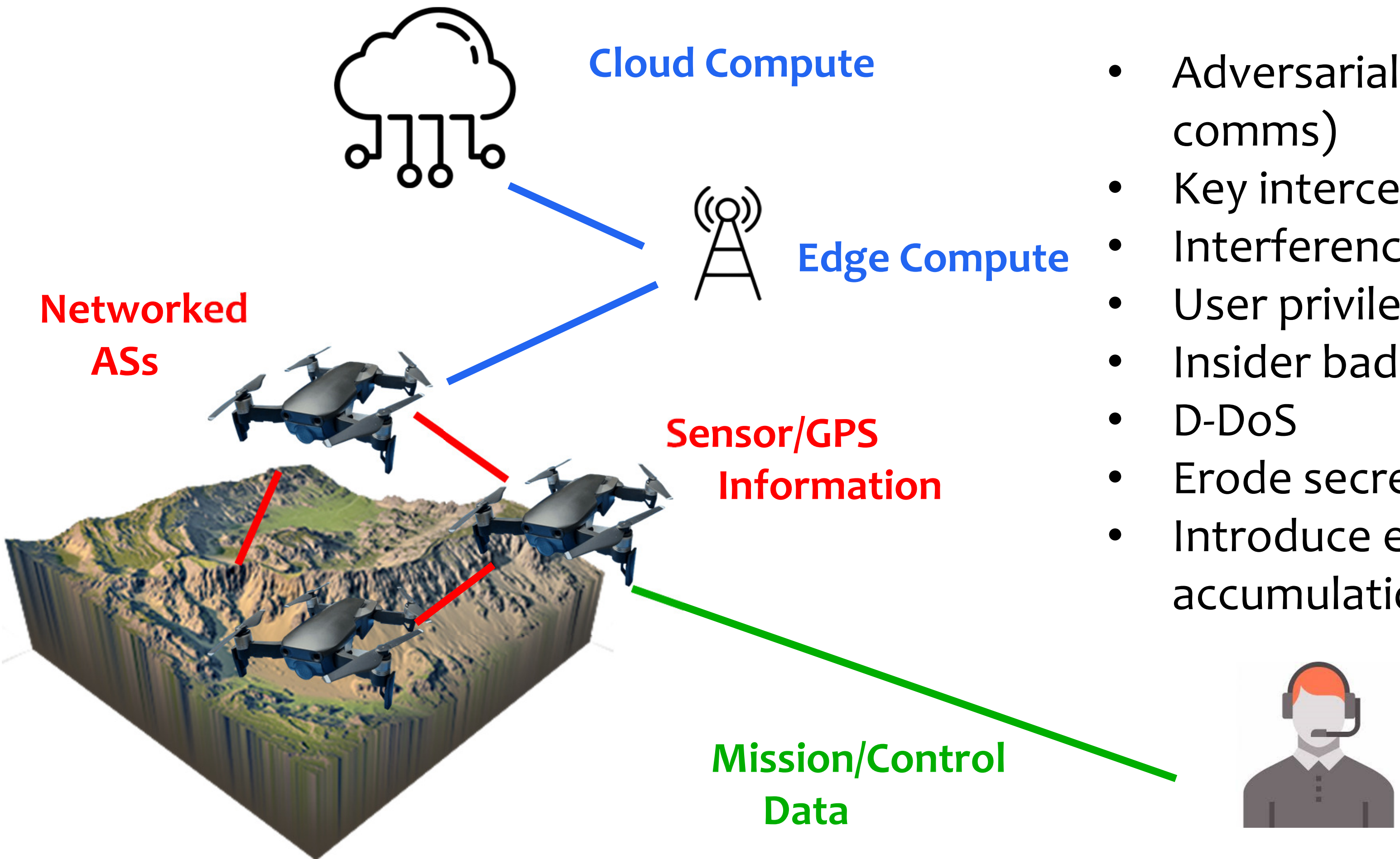
Prof. David Hutchison [Network Security]

Operation Space



Basic Research → Applied → Testbed Validation

Attack Vectors from Physical Signals to Network Packets to Federated Intelligence



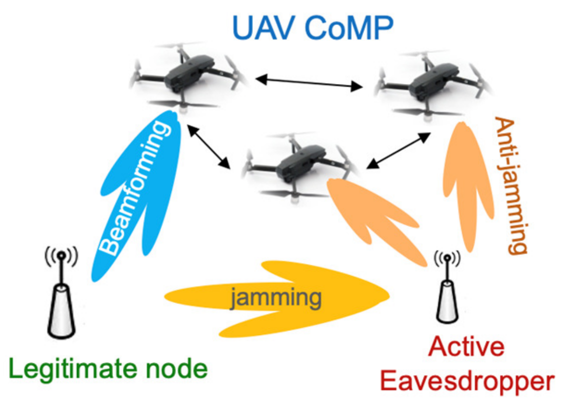
Attack Vectors:

- Adversarial data (sensors, comms)
- Key intercept
- Interference / Jam
- User privilege
- Insider bad behaviour
- D-DoS
- Erode secrecy rate
- Introduce error accumulation

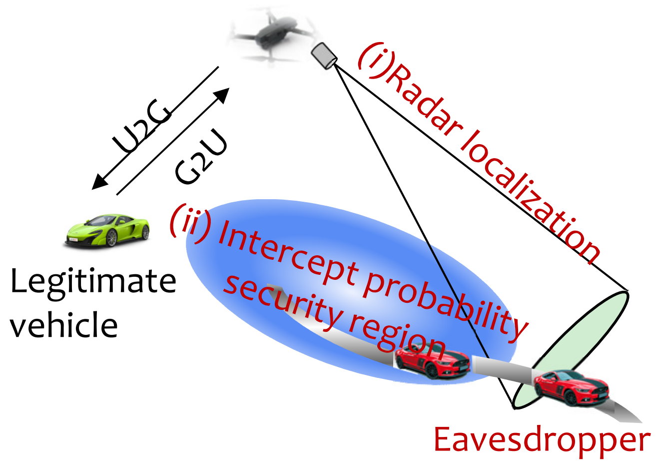
RS-2C – Physical Information Security

Physical Security Review

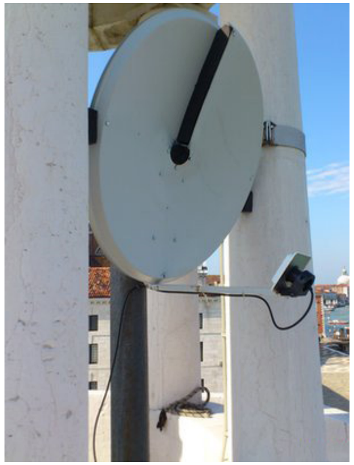
- <u>Purpose</u>: avoid eavesdropper / intercepts through signal shaping
- Attack Vectors: Passive eavesdropping, cooperative active eavesdropping
- Attack type depends on position information of legitimate AS node
- Many physical security techniques out there on beamforming and transmission augmentation



Cooperative beamforming



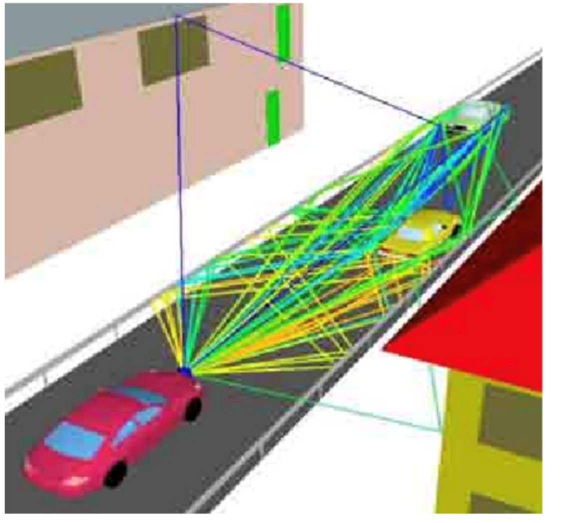
Location Assisted Avoidance



Distortion Modulation

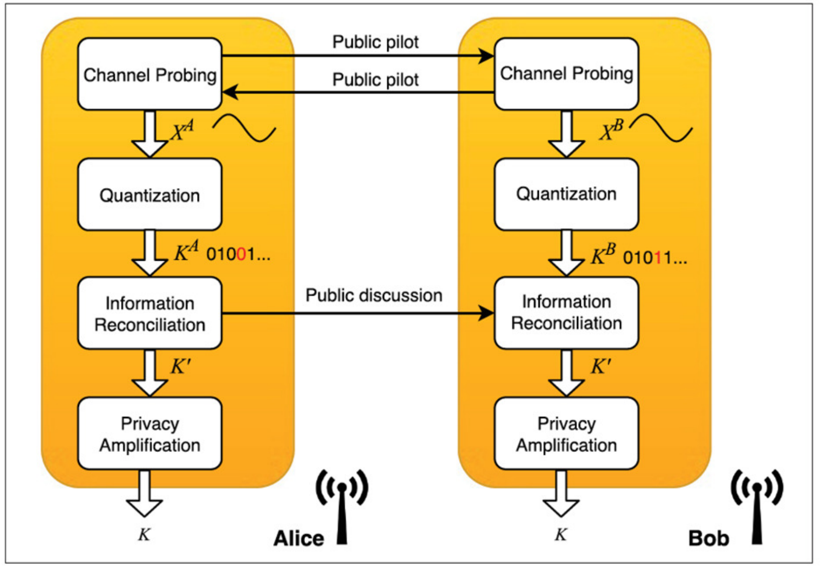
Physical Layer Security: Keys from Mutual Radio Environment

- Purpose: achieve 0 key exchange security at physical signal layer
- Innovation: exploit unique, dynamic, correlated signal features between entities due to the nature of radio signal propagation



Cryptography

- Complex key generation & management & distribution
- No secrecy guaranteed by brute force
- High computational complexity & latency

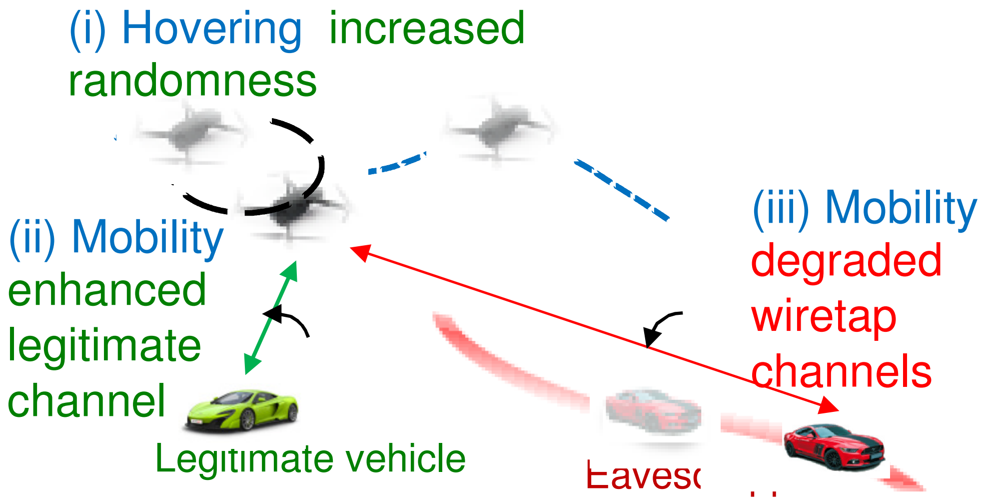


Physical Layer Security (PLS)

Low latency & complexity, key-less, using physical channel properties:
- Randomness of wireless channel
- Superiority of legitimate over wiretap channels

Physical Layer Security: Challenges for Dynamic ASs

Autonomous System improved PLS



Mobility induced Challenges

- Estimate swift time-varying CSIs
- Analyze mobile & silent eavesdroppers
- Optimize secrecy rate with time-varying CSI
- Realistic & complex channel modelling

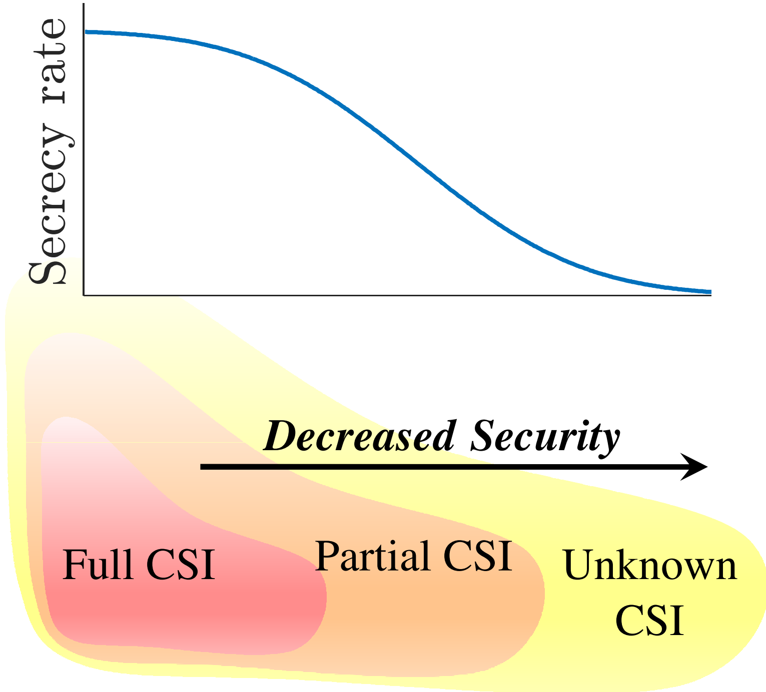
RS-2C – Physical Information Security

Physical Layer Security: CSI Estimation Drives Secure Capacity

- **Full CSI Scenarios:**
  Trajectory/Power/Beamforming optimization [1]
  Precise CoMP anti-jamming & beamforming [2]

- **Partial CSI Scenarios:**
  Robust (worst case) optimization [3]

- **Unknown CSI Scenarios:**
  Friendly Jammer in intercept probability security region [4]



[1] G. Zhang, et al, "Securing UAV Communications via Joint Trajectory and Power Control," in IEEE Trans. Wireless Commun., vol. 18, no. 2, pp. 1376-1389, Feb. 2019.
[2] A. Li, et al, "UAV-Enabled Cooperative Jamming for Improving Secrecy of Ground Wiretap Channel," IEEE Wireless Commun. Lett., vol. 8, no. 1, Feb. 2019, pp. 181–84.
[3] M. Cui et al., "Robust Trajectory and Transmit Power Design for Secure UAV Communications," IEEE Trans. Vehic. Tech., vol. 67, no. 9, Sept. 2018, pp. 9042–46.
[4] Y. Zhou et al., "Improving Physical Layer Security via a UAV Friendly Jammer for Unknown Eavesdropper Location," IEEE Trans. Vehic. Tech., vol. 67, no. 11, Nov. 2018, pp. 11,280–84.

Physical Layer Security: Plan Going Forwards

Fast CSI estimation algorithms for PLS on autonomous vehicles

- CSI estimation for legitimate channels
- Eavesdropper space analysis

Phase 1

Sequential optimization with time-varying CSI

Expand to realistic scenario autonomous systems

Phase 2

Hardware programming & test & realisations

Demonstrate impact with stakeholders

Phase 3

Network Layer Security: Issues with Centralized IP-based Communications to Bypass

- The reliance on the Internet Protocol (IP) opens an array of security risks and vulnerabilities in legacy communication protocols
  - Routing misdirection attacks
  - DNS poisoning
  - IP spoofing
  - Compromised certification authorities

- Routing and topological inefficiencies caused by the centralization and consolidation of IP resources
  - Dependence on cloud providers for data persistence and processing
  - Dependence on CDNs for efficient data delivery
  - Centrally-provided logic on configuration of firewalls, IDS, middleboxes

RS-2C – Network Layer Security

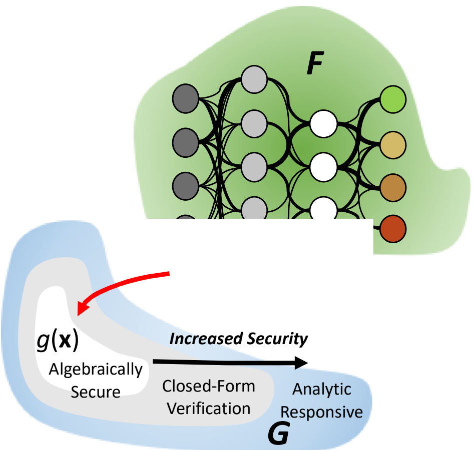
Network Layer Security: Distributed Publish/Subscribe Information Centric Communications

- Information-centric communication model to enable self-organized and dynamic topologies without depending on IP-related resource allocation
  - Structured peer-to-peer network based on a Distributed Hash Table (DHT)
  - Self-organize to exchange disseminate locally observed threats to construct a dynamically updated threat intelligence distributed database

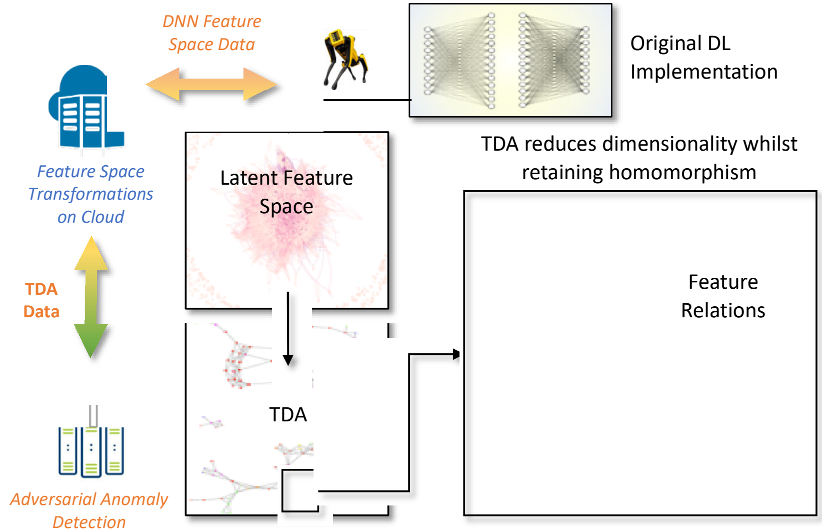
- AS nodes communicate in a publish/subscribe fashion
  - The publish/subscribe protocol combines gossip and epidemic spreading to prevent excessive traffic while also ensure the timely dissemination of messages.

Federated Intelligence Security: Distributed Defence

- **<u>Purpose</u>: Recognise comms and compute are distributed across a networked AS ecosystem. Secure communications in highly complex networked optimisation settings (connect to RS-1B)**
- **Crucial for real-time solutions with multiple KPIs to satisfy. Avoid heuristic optimisation using deep learning**
- **Intelligence is often federated in networked systems (adversarial attacks and defence can occur across communication channels)**
- **<u>Innovation:</u> Develop explainable insight using deep GP, algebraic topology, hypergeometric symbolic, random sketch representations**



Interpretable deep learning with verifiable proofs, adversarial feature detection….etc.

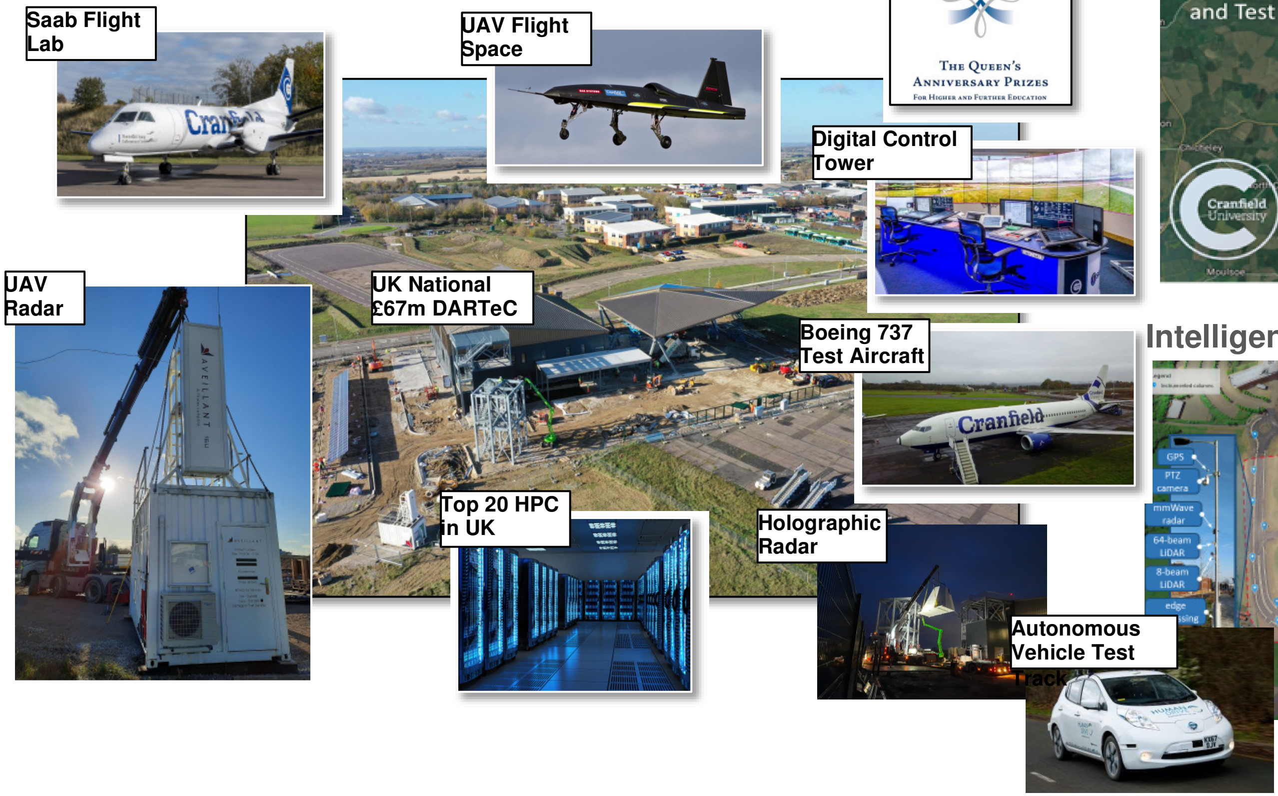


"Scalable Partial Explainability in Neural Networks via Flexible Activation Functions," S. Sun, C. Li, Z. Wei, A. Tsourdos, W. Guo, AAAI Conference on Artificial Intelligence, Feb 2021
"Approximate Symbolic Explanation for Neural Network Enabled Water-Filling Power Allocation" S. Sun, W. Guo, IEEE Vehicular Technology Conference, Apr 2020
"Random Sketch Learning for Deep Neural Networks in Edge Computing," B. Li, P. Chen, H. Liu, W. Guo et al., Nature Computational Science, to appear Feb 2021

Real Autonomous System Test Capability (Theory to Practice)

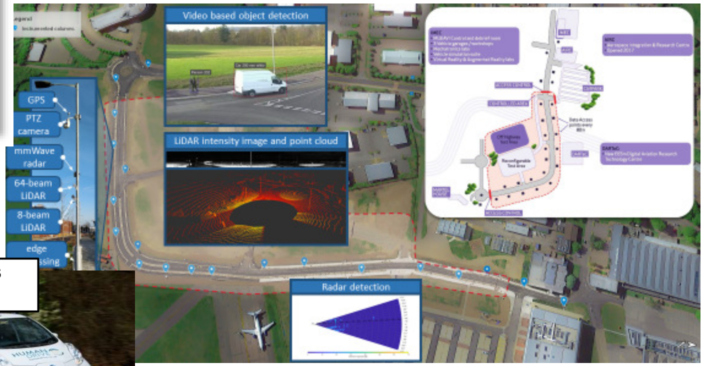
Global Research Airport & Airspace (only 1 in world) with Queen's Award UK flying laboratory

THE QUEEN'S ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION



UK National Unmanned BVLOS Drone Corridor

Intelligent Air-Ground Joint Autonomy Testing



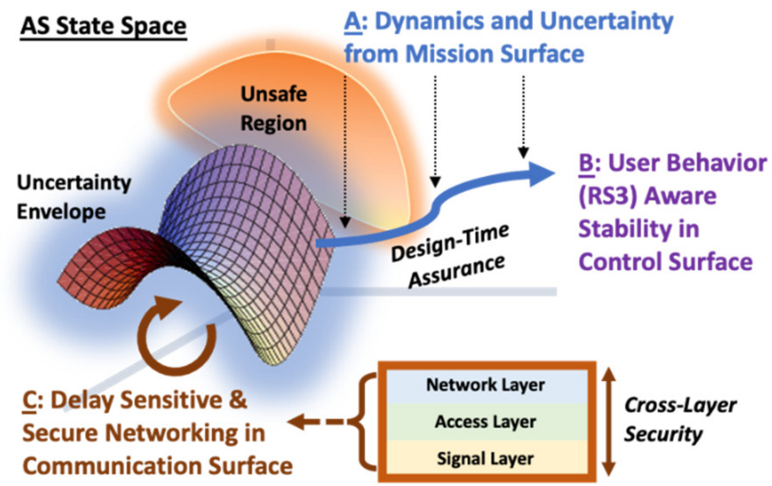
RS-2 Summary

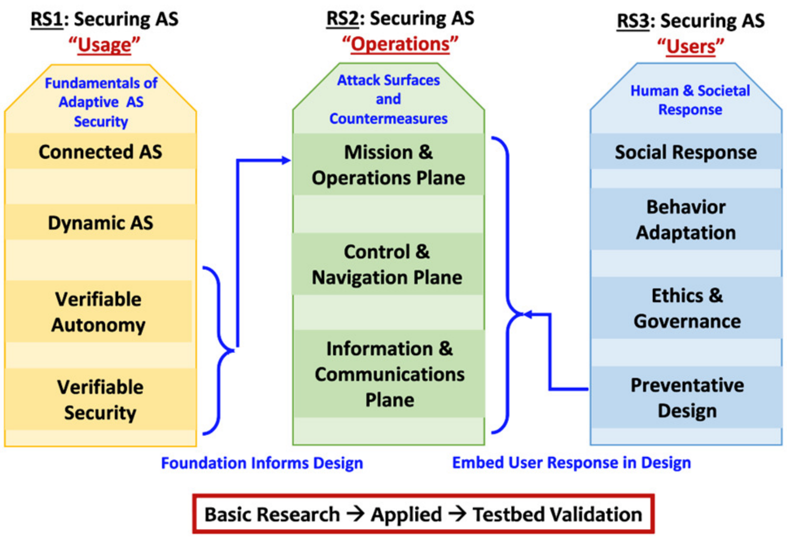
Summary

A. Exposure to cyber-physical attacks by characterizing the attack surfaces, i.e., entry points and likelihoods across the mission surface in a technology & mission-invariant manner.

B. Provide quantifiable safety and feedback to the mission surface when the limits of secure controllability are compromised within a time horizon under current policies and adversarial situations.

C. Provide secure communications across the different layers in the informatics plane from detection of signals to networking.





Research Strand 3

Securing the Autonomous System
“User” Environment

TAS-S Research Strands (RS)

RS1: Securing AS "Usage"

Fundamentals of Adaptive AS Security

- Connected AS
- Dynamic AS
- Verifiable Autonomy
- Verifiable Security

RS2: Securing AS "Operations"

Attack Surfaces and Countermeasures

- Mission & Operations Plane
- Control & Navigation Plane
- Information & Communications Plane

RS3: Securing AS "Users"

Human & Societal Response

- Behavior Adaptation
- Organisational Adaptation
- Ethics, Law & Governance
- Preventative Design

Foundation Informs Design

Embed User Response in Design

Basic Research → Applied → Testbed Validation

TAS-S RS3 Team

Joe Deville

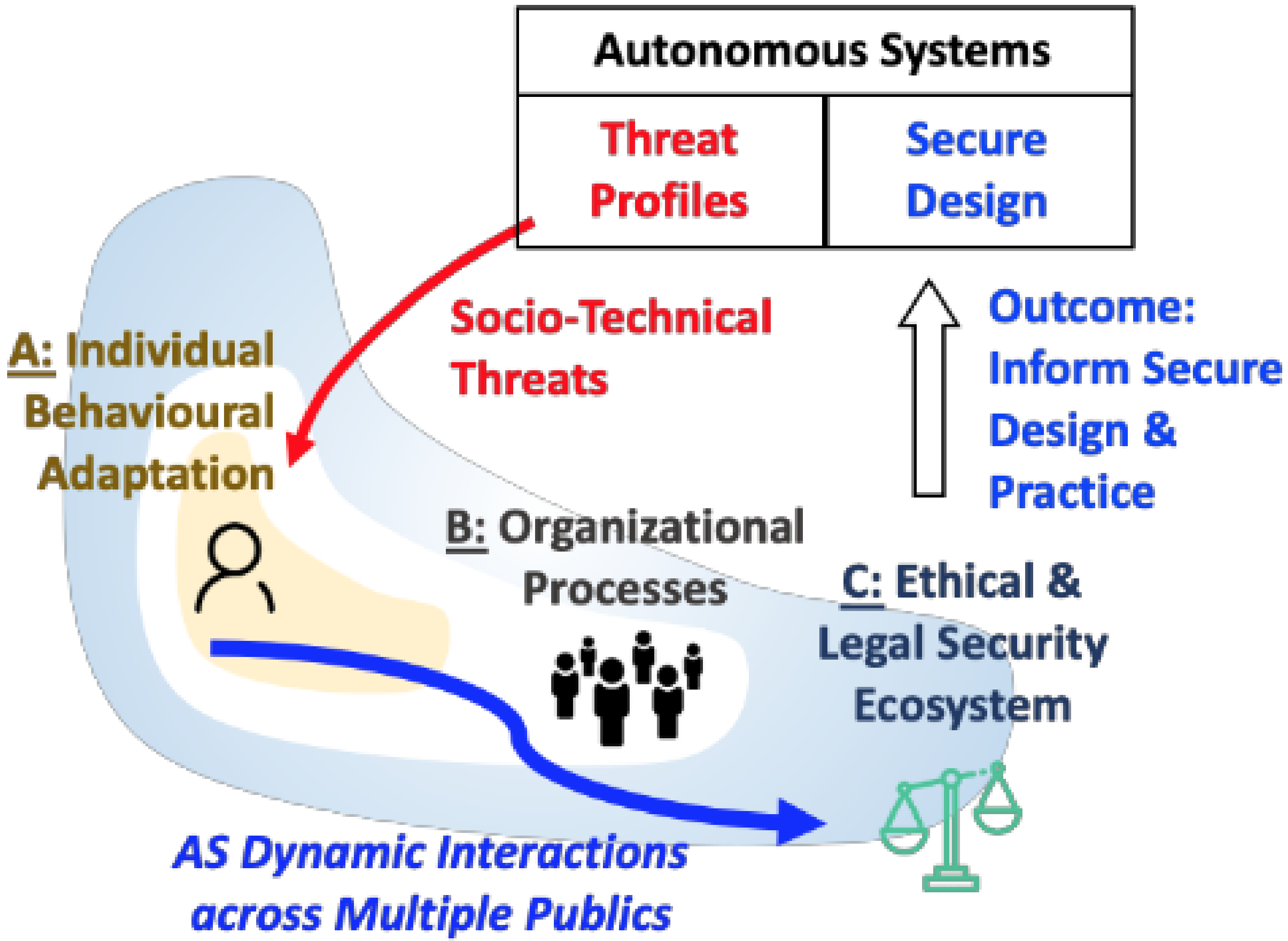
Corinne May-Chahal

Catherine Easton

Lisa Dorn

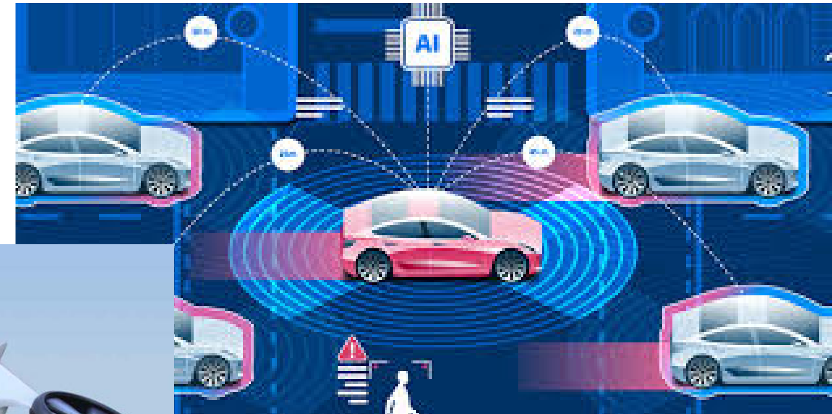
Luke Moffat

Securing the AS “User” environment



AS Scenarios: Discrete, Hybrid, Tethered, Clustered…

Attack Surfaces: Technology + Usage + User Environment



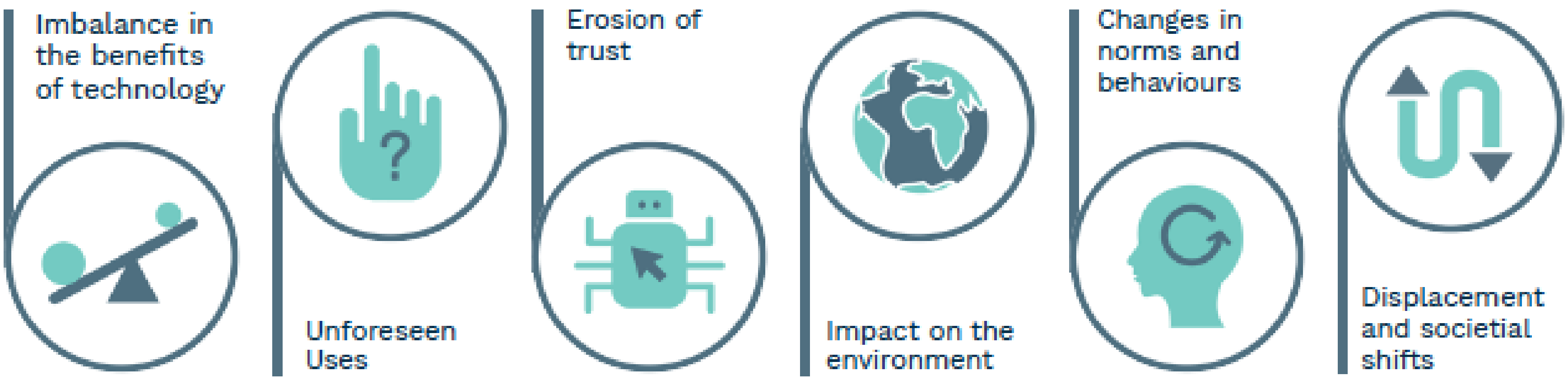
Beginning with an Evidence Informed Approach

Phase 1: What are the human, ethical, legal, social and environmental factors influencing autonomous systems security that have already been researched?

RQ1: What are the human behaviours influencing autonomous systems security?
RQ2: What are the ethical and legal factors influencing autonomous systems security?
RQ3: What are the social factors influencing autonomous systems security?
RQ4: What are the environmental factors influencing autonomous systems security?

Scoping Current Tools and Methodologies

There are many examples of tools and methods that can be adapted to help identify the human, organisational, ethical and social aspects of secure autonomous systems: e.g. TechTransformed resources for Consequence Scanning



Which tools are best adapted to autonomous systems socio-technical security?

TAS-S Research Strands (RS) -

RS1: Securing AS "Usage"

Fundamentals of Adaptive AS Security

- Connected AS
- Dynamic AS
- Verifiable Autonomy
- Verifiable Security

RS2: Securing AS "Operations"

Attack Surfaces and Countermeasures

- Mission & Operations Plane
- Control & Navigation Plane
- Information & Communications Plane

RS3: Securing AS "Users"

Human & Societal Response

- Behavior Adaptation
- Organisational Adaptation
- Ethics, Law & Governance
- Preventative Design

Foundation Informs Design

Embed User Response in Design

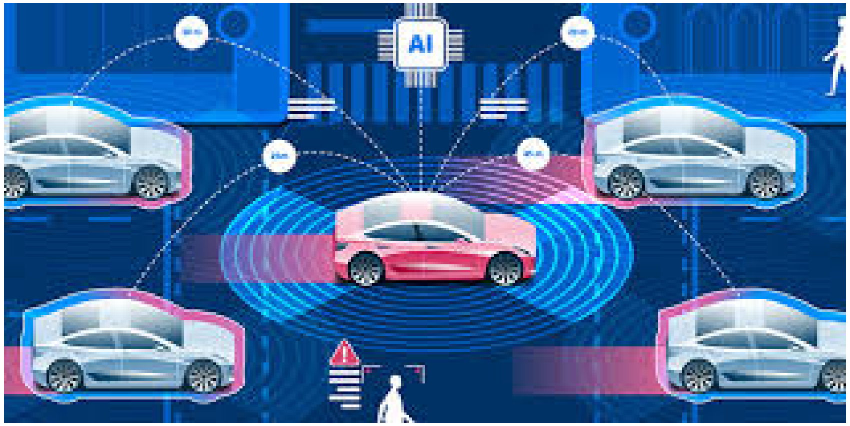
Basic Research → Applied → Testbed Validation

Behavioural adaptation as a basis of security by design

Lisa Dorn
Cranfield University

Behavior adaptation as a security problem

Previous studies to evaluate behavioural adaptation (BA) have been short-term and it is unclear how repeated longitudinal exposure to AS may impact individual response to security threats and threats to the security of others, including the AS itself.



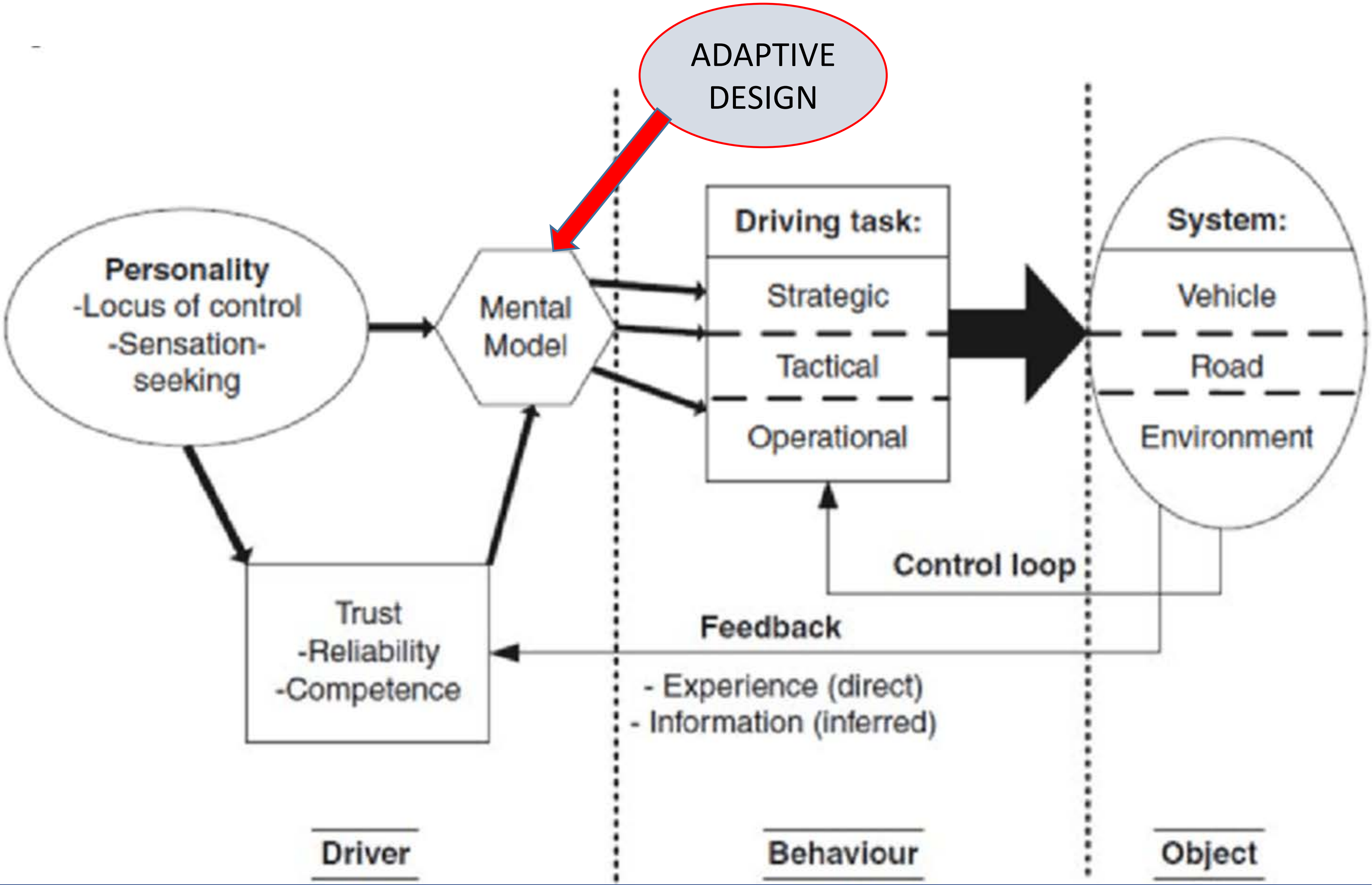
Introduction

- AS design may assume homogenous and static end user behaviour
- Behaviour may change (or adapt) in response to function and performance of an AS - beginning with AVs
- Adaptations may diminish safety and security
- Previous studies are narrow and lab-based

Behavioural Adaptation and AS

Human Factors	Possible BA
Situational Awareness	Recognising that the system is under attack
Monitoring	Ignore alerts; startle response
Workload	Greater secondary task engagement
Trust	As trust increases, attention to critical information decreases
Impairment	Operating the AS whilst impaired (alcohol, drugs, fatigue etc)

Qualitative Model of Behavioural Adaptation



Research Questions

Stage 1: REA - AS's socio-technical security research, adaptive behaviours and regulatory context

- Mental models and how they guide human interaction with AS
- What specific behaviours change as humans adapt to AS and how might this change compromise security?
- Previous experience with technology and BA to AS
- Measurement of BA

TAS-S Research Strands (RS) -

RS1: Securing AS "Usage"

Fundamentals of Adaptive AS Security

- Connected AS
- Dynamic AS
- Verifiable Autonomy
- Verifiable Security

RS2: Securing AS "Operations"

Attack Surfaces and Countermeasures

- Mission & Operations Plane
- Control & Navigation Plane
- Information & Communications Plane

RS3: Securing AS "Users"

Human & Societal Response

- Behavior Adaptation
- Organisational Adaptation
- Ethics, Law & Governance
- Preventative Design

Foundation Informs Design

Embed User Response in Design

Basic Research → Applied → Testbed Validation

Organisational adaptation as a security problem

Cultures and practices within organizational settings can also dramatically change the design and success of secure AS. Secure AS require new methodologies that organisations can use to critically interrogate socio-technical processes and their engagement with wider publics.

TAS-S Research Strands (RS) -

RS1: Securing AS "Usage"

Fundamentals of Adaptive AS Security

- Connected AS
- Dynamic AS
- Verifiable Autonomy
- Verifiable Security

RS2: Securing AS "Operations"

Attack Surfaces and Countermeasures

- Mission & Operations Plane
- Control & Navigation Plane
- Information & Communications Plane

RS3: Securing AS "Users"

Human & Societal Response

- Behavior Adaptation
- Organisational Adaptation
- Ethics, Law & Governance
- Preventative Design

Foundation Informs Design

Embed User Response in Design

Basic Research → Applied → Testbed Validation

Ethics, law and governance as a security problem

Ethical, Legal and Social Issues (ELSI) surrounding AS security interact with a wide range of overlapping aspects of AS development. These are constantly changing as AS evolve. How can designers and developers best adapt?

New methods for the design of more ethical, more secure Autonomous Systems

Joe Deville, Catherine Easton, Luke Moffat
Lancaster University

Knowing ethics, law & security in relation to AS's

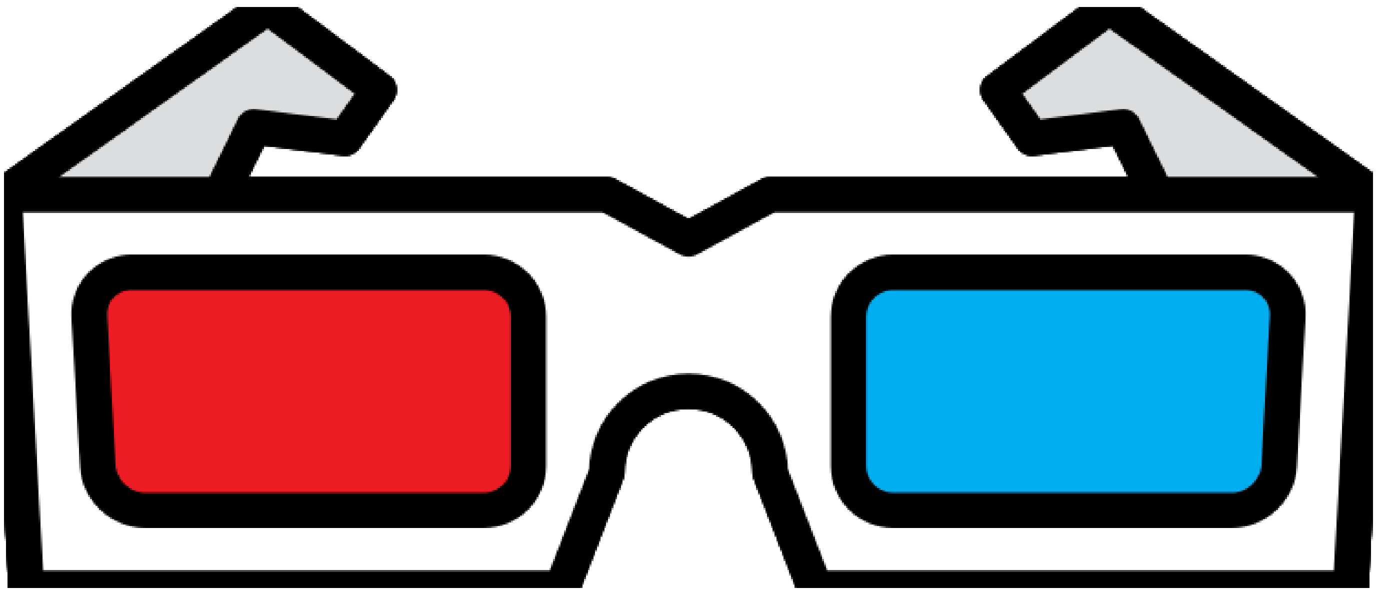
Developing a new method for addressing core project RQs:

- What are the ethical and legal factors influencing autonomous systems security? (RQ2)
- What are the social factors influencing autonomous systems security? (RQ3)
- What are the environmental factors influencing autonomous systems security? (RQ4)

Assumes questions of ethics and security are not reliably knowable in advance of interactions between technologies and society (=> uncertainties/unknown unknowns)

- We need to understand how AS security, law and ethics are understood by diverse stakeholders – wider contemporary views on ethics & security
- We need to understand how diverse stakeholders imagine their likely and desirable futures with AS's – wider future-focused views of ethics & security
- We need to provide actionable guidance to organisations looking to develop more secure/ethical AS's & to deal with the complexity of the social

A wider view on ethics & security

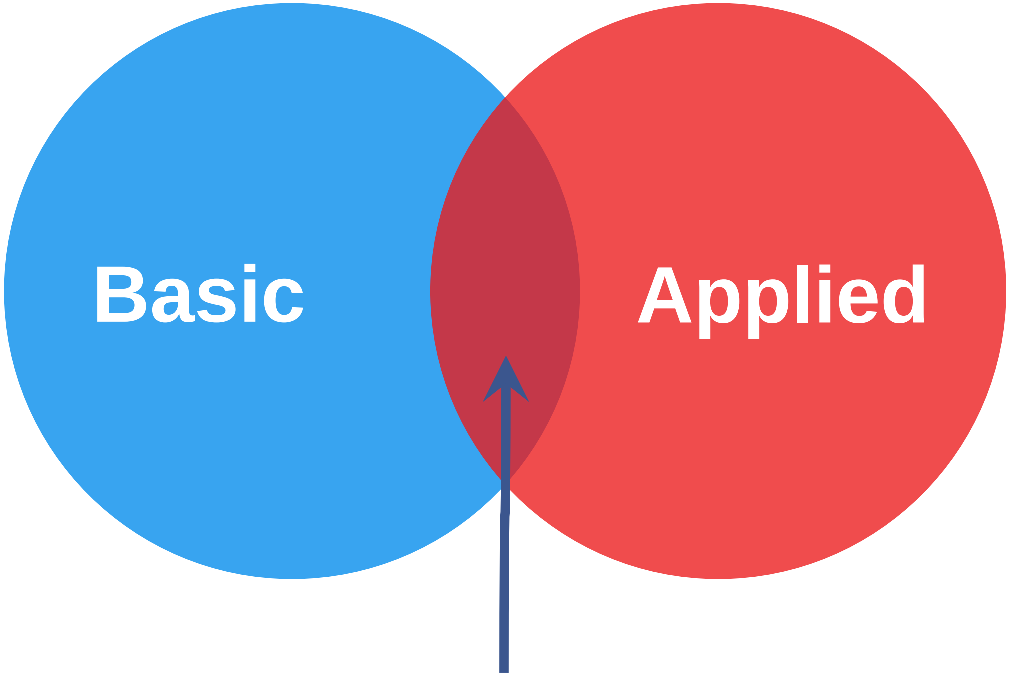


A wider view on ethics & security

Systems

Actors

Accountabilities



Engagement

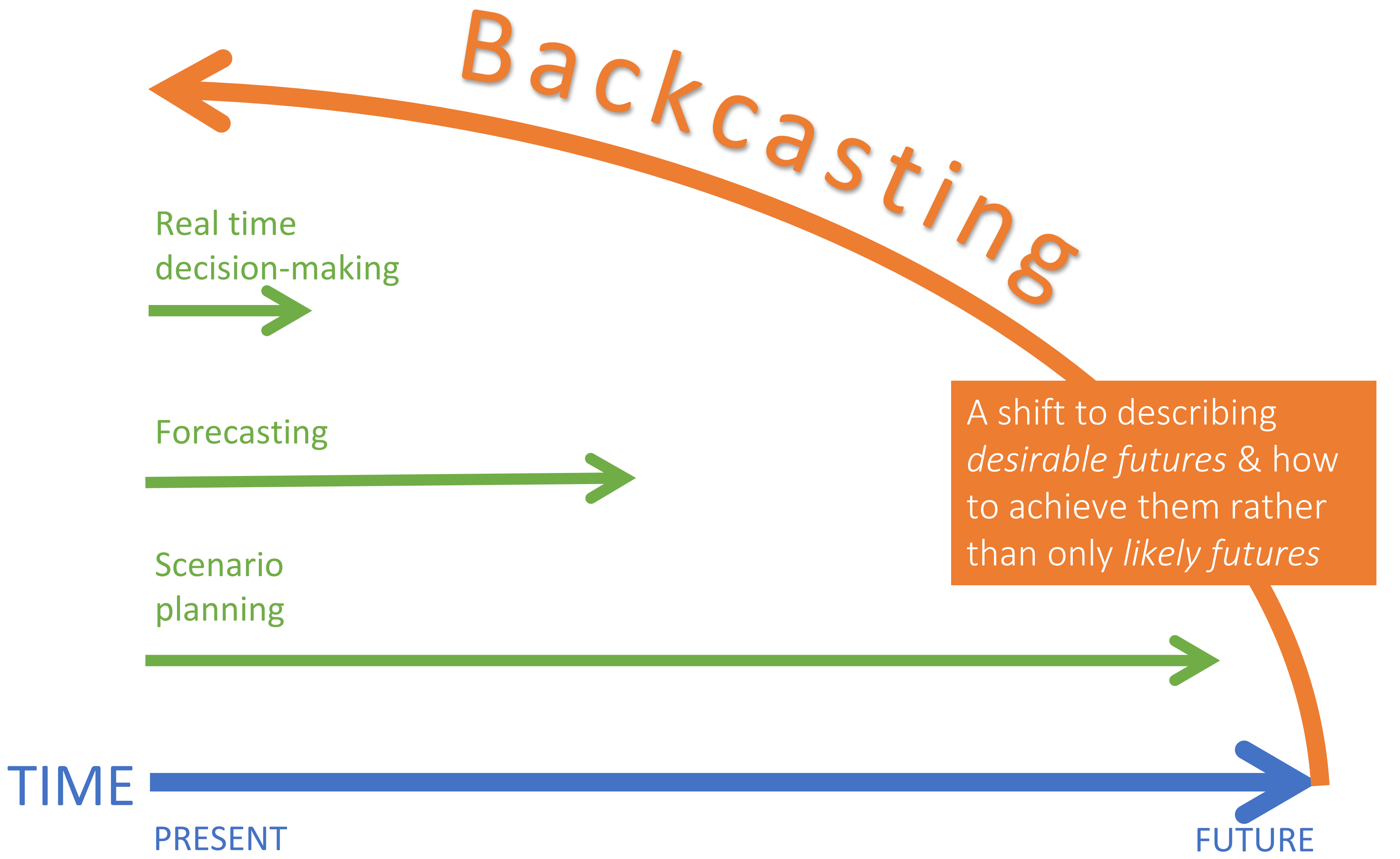
Participation

Exploring ethics & security through co-design

A wider view on law & regulation

- On-going, iterative conversation: this is not simply a story of compliance
- Engagement with hard law (legislation eg GDPR) and soft, regulatory measures (eg certification)
- Opportunity to feedback on the substantive practices that embed legal issues into collaborative technology development
- ELSI methods will draw out the interaction and interplay between law and ethics
- On a wider level industry standards and co-regulatory security measures will be analysed, through engagement with stakeholders

Scoping present & future interactions with technologies



Combining backcasting, controversy analysis & ELSI

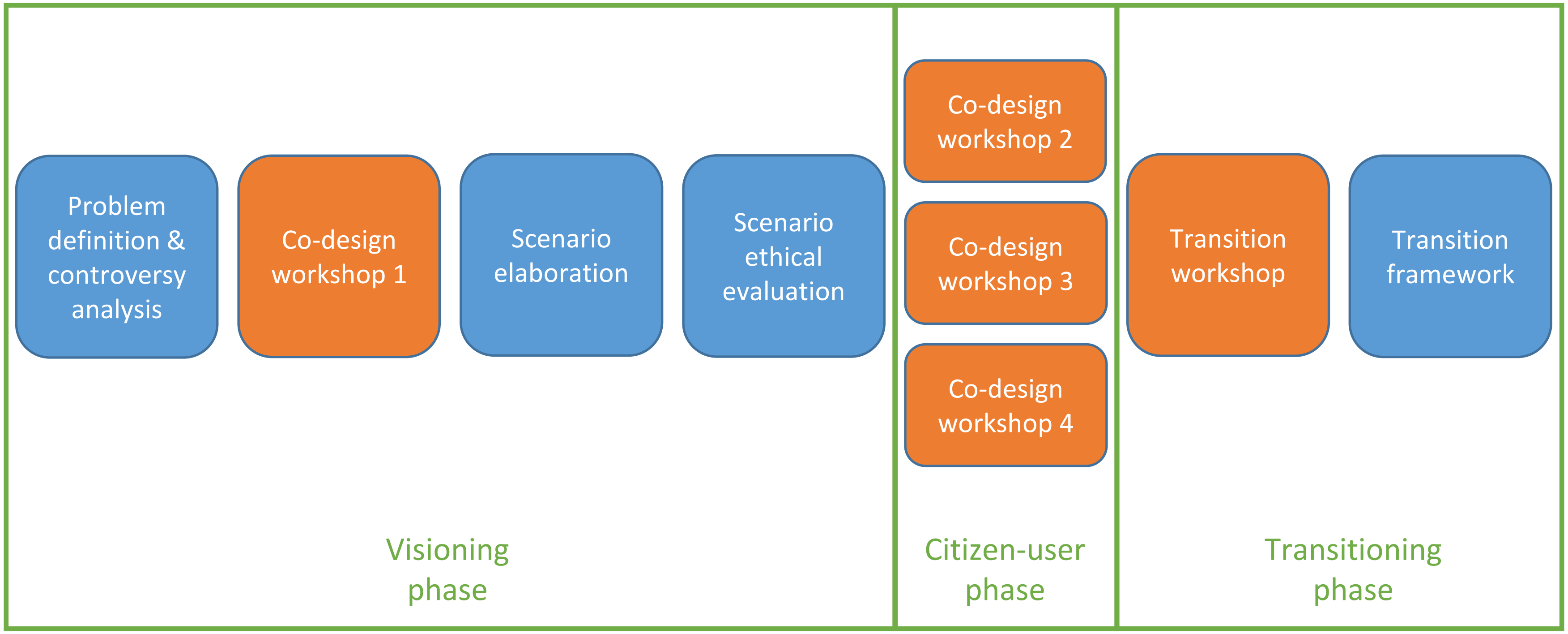
Participatory backcasting: emphasis on using engagement with diverse affected parties/stakeholders as a resource to broaden perspective on issues characterised by uncertainty

Controversy analysis: interested in understanding which groups ('publics') are shaping debates around new technologies, and analysing the assumptions brought to engaging with a new technology by these groups

ELSI: interested in understanding the interplay between ethics and law and its wider impacts on society. Looks both internally at the project and externally through stakeholder engagement

In all three approaches, seen as vital to use both expert and lay forms of knowledge as resources for understanding the unknowns and ethical issues surrounding a new social/technological developments

Informing the design of more ethical, secure ASs

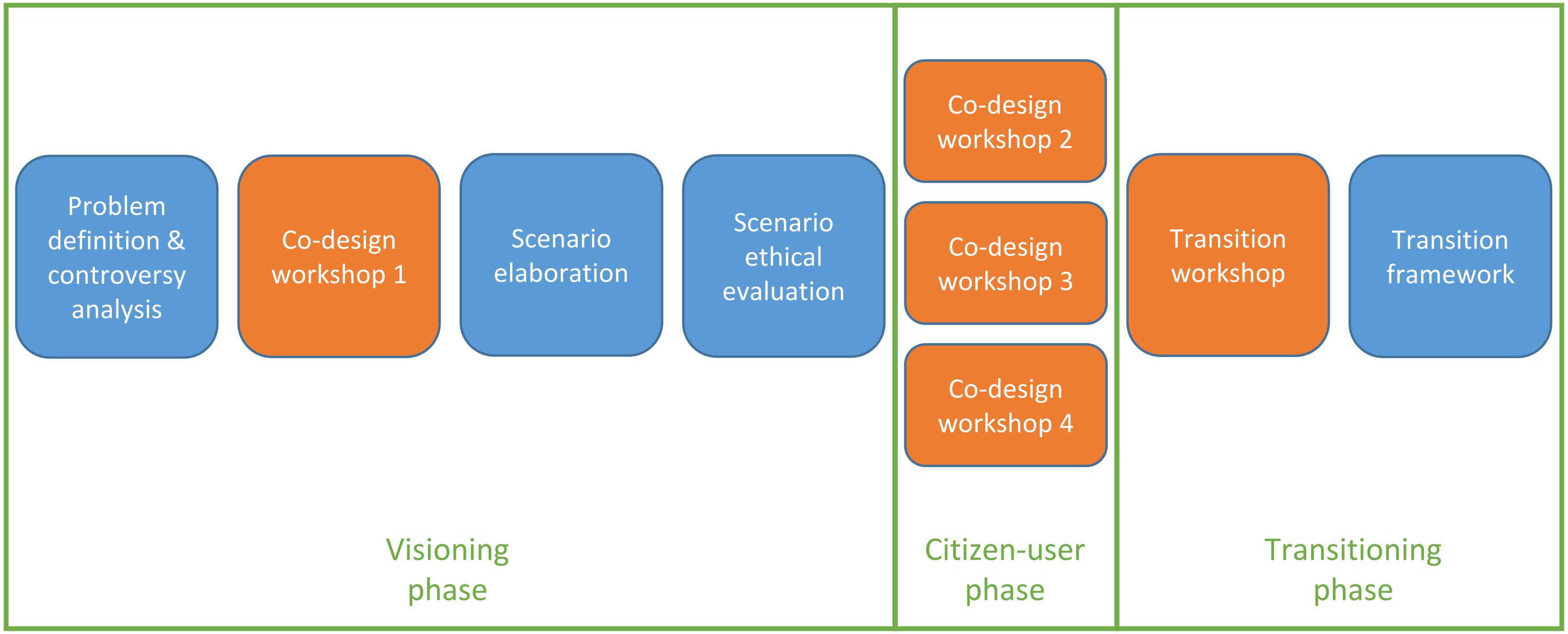


Combining ELSI & Controversy Analysis within a participatory backcasting framework to identify key stakeholders/publics, contemporary ethical & security issues, and future-focused visions for secure, ethical AS's

Elaborating visions, assessing feasibility

Working with partners to identify practical interventions

Informing the design of more ethical, secure ASs



Using this process to work in depth with 2 case study partner organisations before developing a set of resources – including a best practices, a handbook, multimedia resources – for use by organisations looking to develop secure, ethical, autonomous systems

Using co-design for knowledge exchange

- Supports and drives industry R&D capacity for socio-technological innovation, in response to ELSI, via knowledge exchange, creative and participatory methods.

- Co-designs spaces and tools with which designers, engineers, practitioners, communities and policymakers, collectively consider and anticipate better futures.

- By co-designing capacity, we mean, opening spaces, building frameworks, and creating tools to make technological development response-able.

- ‘Ethics through Design’ as a framework for co-designing security.

isITethical? Values

AUTONOMY

BENEFICENCE

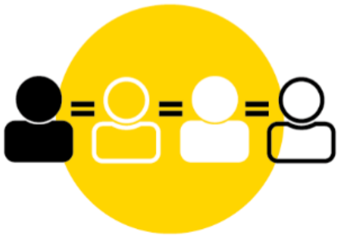
CO-OPERATION

CONSENT

DATA PROTECTION

DIGNITY

DIVERSITY



EQUALITY

Today's workshop

What is the objective?	Discover our encounters with ethics
What is the value?	Getting to know each other, finding common ground and differences
What are the requirements?	45 mins Group work in breakout rooms

Next steps

Scoping legal & regulatory issues with key stakeholders

- Work to begin immediately following this introductory workshop
- What are the legal and regulatory issues that are challenging for you in relation to ethics & security?
- How can the project contribute to debates on changes to legal structures (esp. post Brexit)
- 1:1 scoping interviews – your contributions v. much needed & appreciated

To inform the design of a future stakeholder workshop

- Focusing on ethical and social issues associated with autonomous systems security intersect (or not) with legal issues

Getting Involved

Review Briefings

Backcasting and controversy analysis

Tool development