ifip Session 4: Assessment and Defense Methods WG.10.4

- **How to defend connected intelligent vehicles: Transferring established Information Security best practices to the vehicular world**
  Miriam Gruber and Jan Lange, Volkswagen

- **Detection is not enough: Low-cost Attack Recovery for Autonomous Robotic Vehicles (RVs)**
  Karthik Pattabiraman, University of British Columbia

- **Design and Assessment of Safe Autonomous Vehicles (AVs)**
  Saurabh Jha, IBM T. J. Watson Research

- ***Session Chair:*** *Andrea Ceccarelli, University of Florence*
- ***Rapporteur:*** *Homa Alemzadeh, University of Virginia*

Transferring IT Security Best Practices to IVs

WG.10.4

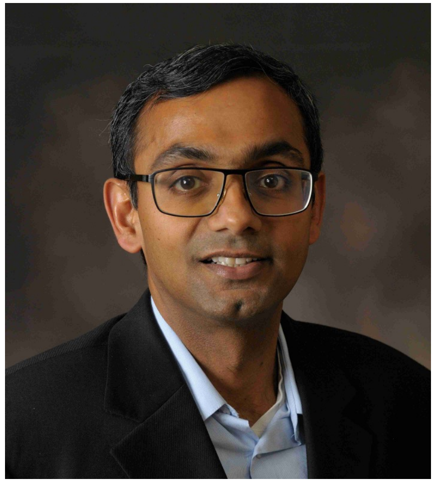
- **Collision of two worlds:** Information security and automotive safety
- **From prevention to active defense**
  - **Prevention:** Interface protection, SW integrity, authenticated communication
  - **Defense:** Intrusion detection, intrusion reaction, active defense, and recovery
  - **Challenges:**
    - New technology, timing constraints, increasing complexity, fixed rules
- **Incident response:**
  - **Active Attack Detection**
  - **Response**
- **AI-based defense:**
  - **AI-based Detection:**
    - Learn from real-world attack scenarios, not enough data
  - **AI-based Response:**
    - Too risky, needs absolute certainty, not enough real-world data to train on

Transferring IT Security Best Practices to IVs

WG.10.4

- **Active Attack Detection**
  - **Steps:**
    - **Vehicle:** Collect data from vehicle **=>** Apply anomaly detection rules
    - **Backend (Cloud):** Aggregate data (fleet-wide) **=>** More in-depth detection
  - **Challenges:**
    - **What data?** Data from ECUs, interfaces (e.g., Wifi, Bluetooth), V2V communications
    - **How much data?** Just enough to analyze the attacks and the infrastructure
  - **Best practices:**
    - Asset register (ECUs), asset use cases, review by service owners
- **Response**
  - **Goals:**
    - Contain or mitigate attacks **=>** Stop incident **=>** Recover **=>** Lessons learned
  - **Challenges:**
    - Variety of attack models with different levels of intelligence and complexity
  - **Best practices:**
    - Safety-critical context/usage, context-specific fall-back, automated vs. manual response

Low-cost Attack Recovery for RVs

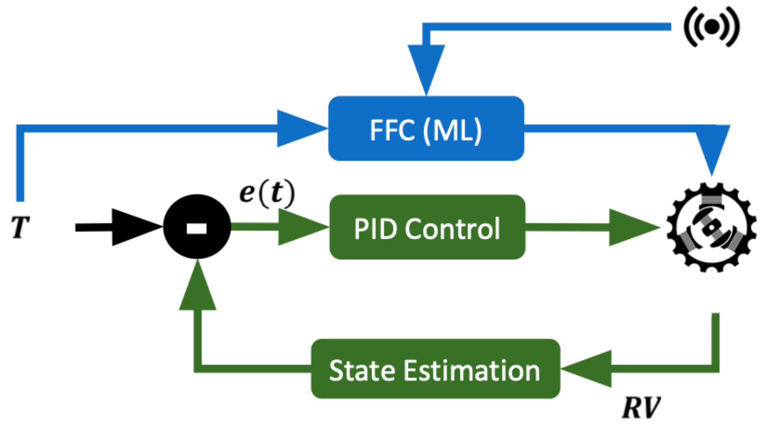


- **Perception in RVs**
  - Sensor attacks
  - Can RVs continue to operate safely despite sensor attacks?
- **State-of-the-art Attack Detection and Recovery**
  - **Detection:** Invariant-based and model-based
  - **Recovery:** Fail-safe mechanisms (emergency landing)
- **Attack Recovery without mission failure or crash**
  - Prevent erroneous physical states AND prevent erroneous actuator signals
  - **PID-Piper**
    - **Problem:** PID overcompensation under attacks => good for faults, not for attacks
    - **Solution:** Redundant feed-forward controller (FFC)
  - **DeLorean**
    - **Problem:** Multiple sensors under attack
    - **Solution:** Identify attacked sensors, isolate them, substitute sequence, recover by replay

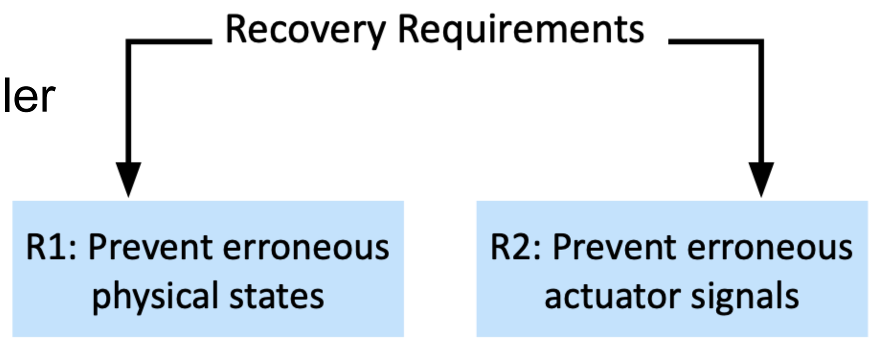
Low-cost Attack Recovery for RVs

WG.10.4

- **Attack Recovery Methods**
  - **PID-Piper**
    - Redundant feed-forward controller to address PID overcompensation
    - ML trained on sensor and waypoint data to predict recovery actions
    - Switched to upon attack detection and active for the attack duration
    - **Higher mission success, low false positives, negligible overhead**



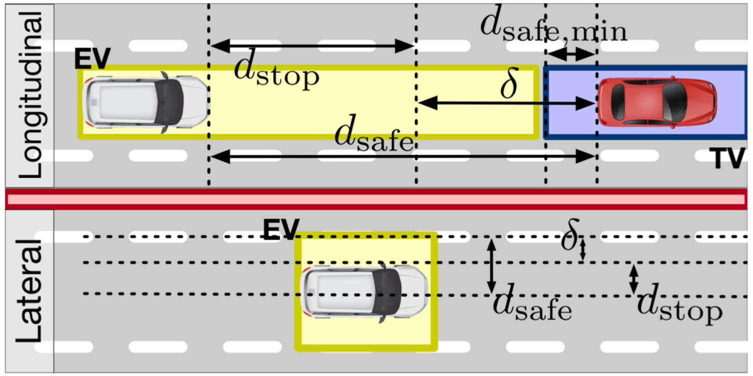
  - **DeLorean**
    - Detect the attacked sensors
    - Prevent erroneous physical states: isolating sensor(s) from controller
    - Prevent erroneous actuator signals: substituting input sequence
    - Discard corrupted states and replay historic states
    - **First work to recover from multiple sensor attacks with little overhead**



Design and Assessment for AV Safety

WG.10.4

- **Vulnerabilities in AVs**
  - Much worse than non-AVs
  - Increased attack surface: ML uncertainty, training data quality, unknown unknowns
- **Identifying safety-critical vulnerabilities**
  - **Problem:** State-space exploration to find the faults that lead to safety hazards
  - **What/Where to inject faults?**
    - **Solution:** Accelerate testing by only doing FI based on ML inference
      - Probabilistic Graph Models (PGMs) to model fault propagation
      - Training on observational data
      - Model fault injection as an inference query on PGM
  - **What/How/When to launch attacks?**
    - **Solution:** Design Ml-driven attacks that can evade detection
      - Alter objects trajectories by corrupting pixels or perception output
      - ML inference of low safety potential and minimum time to hazards
  - **Much faster and more efficient identification of safety-critical scenarios than random FI**
- **Runtime threat assessment for safety**



Session 4: Summary

WG.10.4

- **Current Challenges**
  - IT to AV transfer of security and safety methods and best practices
  - Lack of realistic incident data and labels for training detection and response models
  - Effect of ML uncertainties and quality of training data
  - Timing constraints, computational overhead, and side consequences of methods at runtime

- **Future Directions**
  - ML/AI driven models for fault injection, safety assessment, attack detection and recovery
  - Combined model and data-driven methods, situationally-aware methods, both online and offline
  - Simulation to real transfer of safety models, fault and driving scenarios, and datasets
  - Community standards for quantifying the quality of ML models and datasets