# Multi-Label Machine Learning-Based Identification of Concurrent Cyberattacks in Connected Autonomous Electric Vehicles

Zeyuan Xu
*Department of Electrical, Computer and Biomedical Engineering*
*University of Pavia*
Pavia, Italy
zeyuan.xu@unipv.it

Beiyi Zhang
*Xi'an Institute of Applied Optics*
*China North Industries Group Corporation Limited*
Xi'an, China
zhangbeiyi@sjtu.edu.cn

Liang Zhang
*School of Electrical Engineering and Automation*
*Anhui University*
Hefei, China
liangzhang@ahu.edu.cn

Shixiang Lu
*College of Electrical and Information Engineering*
*Northeast Petroleum University*
Daqing, China
lsx94@nepu.edu.cn

Antonella Ferrara
*Department of Electrical, Computer and Biomedical Engineering*
*University of Pavia*
Pavia, Italy
antonella.ferrara@unipv.it

***Abstract*—Concurrent cyberattacks can produce overlapping motion and communication anomalies in connected autonomous electric vehicles. This paper develops a multi-label overlap-aware separation and identification of cyberattacks (MOSAIC) framework that couples receiver-observable attack labels with grouped physical and communication evidence. Its causal receive-buffer formulation separates attempted attacks from observable effects, avoiding contradictory supervision when concurrent mechanisms mask a transmitted field. Group-specific projections combine nominal-model residuals, motion discrepancies, and field-level communication metadata within a compact recurrent encoder. Attack-specific evidence queries and independent sigmoid heads support simultaneous identification without enumerating attack combinations. An episode-disjoint protocol distinguishes familiar overlaps, unseen combinations, and operating-domain shifts, while ablations assess the contribution of communication evidence and learned aggregation. Comparative tests validate the effectiveness of the proposed framework for concurrent cyberattack identification.**

***Index Terms*—Connected autonomous electric vehicles, concurrent cyberattacks, multi-label identification, recurrent learning, receiver observability.**

## I. Introduction

Connected autonomous electric vehicles (CAEVs) regulate their longitudinal motion using local measurements and information exchanged with neighboring vehicles. A corrupted sensor and a compromised vehicle-to-vehicle (V2V) field may therefore affect the same feedback loop, even when they originate from different attack mechanisms. Their combined effects propagate through torque commands and inter-vehicle spacing, so the received anomaly is not necessarily a simple superposition of isolated signatures. A useful vehicle-side detector should report the mechanisms supported by its available evidence without confusing ordinary communication imperfections with deliberate interference. Distributed learning and control provide tools for interconnected nonlinear systems [1]; reliable monitoring also draws on advances in manufacturing diagnosis [2], redundant inertial sensing [3], and class-imbalanced decentralized learning [4]. For vehicle-side security, the central question is whether a detector can identify the attacks that remain observable when several mechanisms operate together.

Hybrid neural/model-based vehicle monitoring [5] and self-attention sensor analysis [6] offer useful ways to combine temporal evidence. Their original diagnostic settings, however, do not by themselves establish identification of concurrent field-specific attacks. Multi-label learning permits several labels per observation [7], but changing a classifier's output activation cannot resolve a loss of information. The monitoring limitations studied in [8] are relevant here: when denial of service (DoS) blocks a field, an additional modification of that same field may leave no distinguishable received trace. Training against the attacker's intentions alone can consequently impose an unattainable target. Likewise, false-data injection (FDI) and replay may disturb related motion features while requiring different diagnostic evidence. Field-specific update loss and information age can complement physical residuals, provided these indicators are generated causally rather than supplied as ideal attack flags.

This paper develops MOSAIC around three connected el-

This work was supported in part by the European Union's Horizon Europe Program for Marie Skłodowska-Curie Actions (FEASIBLE) under Grant 101149415, and the Key Program of Natural Science Foundation of Anhui Higher Education Institutions of China under Grant 2022AH050068.

ements: a causal receive-buffer simulator that distinguishes attempted from observable attacks; a 29-dimensional, five-group representation coupled to a compact gated recurrent unit (GRU) multi-label classifier; and a protocol that separates seen pairs, unseen higher-order combinations, and operating-domain shifts. The study evaluates complete label-set recovery as well as per-label F1, preventing a strong aggregate score from concealing missed concurrent mechanisms. Comparisons and controlled ablations identify which parts of the representation help, while reporting the cases in which learned group attention does not improve identification. The contribution is therefore an observability-consistent diagnostic formulation and reproducible separation of concurrent evidence, not an assumption that every attempted attack remains identifiable. All reported transfer tests retain the training-selected normalization and decision thresholds.

## II. Description of Connected Autonomous Electric Vehicles

Consider three heterogeneous CAEVs traveling on a flat road, with vehicle 1 leading and vehicle 3 receiving position and velocity fields from vehicles 2 and 1. Fig. 1 focuses on this two-predecessor receiver topology. The simulated longitudinal state is $x_n = [p_n, v_n, T_n]^\mathsf{T}$, where $p_n$, $v_n$, and $T_n$ denote position, velocity, and combined traction/braking torque. Using a torque-lag longitudinal model [9], its forward-Euler dynamics with sampling interval $\Delta t$ are

$$\begin{aligned} p_n(k+1) &= p_n(k) + \Delta t\, v_n(k), \\ v_n(k+1) &= v_n(k) + \frac{\Delta t}{m_n}\left(\frac{\eta_{T,n}}{R_n}T_n(k) - C_{A,n}v_n^2(k) \right. \\ &\qquad \left. - C_{r,n}m_n g\right), \\ T_n(k+1) &= T_n(k) + \frac{\Delta t}{\tau_n}\big(u_n(k) - T_n(k)\big). \end{aligned} \tag{1}$$

Here $m_n$, $R_n$, $\tau_n$, $\eta_{T,n}$, $C_{A,n}$, and $C_{r,n}$ represent mass, wheel radius, torque-response lag, driveline efficiency, aerodynamic drag, and rolling resistance. The electric-vehicle abstraction covers longitudinal torque response; it does not model battery electrochemistry.

The follower controller uses received spacing and speed discrepancies. Define $\epsilon_{p,j} = p_j^{\mathrm{rx}} - \widetilde{p}_3 - (3-j)d_0$ and $\epsilon_{v,j} = v_j^{\mathrm{rx}} - \widetilde{v}_3$, for $j \in \{2,1\}$. Its bounded acceleration request is

$$\begin{aligned} a_3^c = \mathrm{clip}\big(&0.14\epsilon_{p,2} + 0.42\epsilon_{v,2} \\ &+ 0.035\epsilon_{p,1} + 0.12\epsilon_{v,1}, -3.5, 2.5\big). \end{aligned} \tag{2}$$

The torque command follows the inverse force balance in (1), using $a_3^c$ and measured velocity. Corrupted local and received measurements enter this feedback loop, so attacks can change subsequent physical motion without directly overwriting plant states.

Each predecessor's position and velocity have separate receive buffers. For field $f \in \{p, v\}$, let $\ell_{j,f} = 1$ denote a missing update. A received field follows

$$\xi_{j,f}^{\mathrm{rx}}(k) = \begin{cases} \xi_{j,f}^{\mathrm{rx}}(k-1), & \ell_{j,f}(k) = 1, \\ \xi_{j,f}(s_k) + \delta_{j,f}(k), & \ell_{j,f}(k) = 0. \end{cases} \tag{3}$$

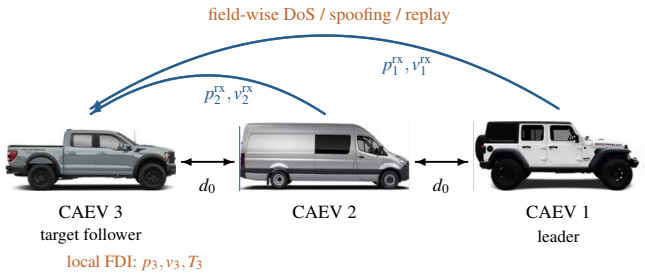


Fig. 1. Receiver-focused two-predecessor topology. Local FDI affects vehicle 3 measurements; remote attacks affect distinct position/velocity receive fields.

The ordinary source lag is $\max(1, \lceil d_{j,f}/\Delta t \rceil)$; replay adds its own source lag before retrieving the value, and $s_k \geq 0$ is clipped at the episode origin. Thus delivered data never uses a future sample. A dropped update holds the previous value and increases its age. Spoofing modifies the claimed sender's position through $\delta_{j,p}$, whereas local FDI corrupts $\widetilde{p}_3$, $\widetilde{v}_3$, or $\widetilde{T}_3$.

The simulated packet format is field-addressable: position-only and velocity-only DoS are meaningful under this assumption. If DoS suppresses a field throughout an interval, modifications of its undelivered payload cannot be reconstructed from that buffer alone. The main dataset therefore allocates concurrent remote mechanisms on the same field to different predecessor links. This is a target-observability restriction, not a security countermeasure, and does not preclude simultaneous local and remote corruption of the same physical quantity.

## III. Intelligent Detection of Concurrent Cyberattacks

### A. Feature and Label of CAEVs

At instant $k$, the detector constructs $z(k) \in \mathbb{R}^{29}$ from five ordered groups. The physical group contains the three local measurements, $u_3$, and the four received predecessor states. Three one-step residuals compare local measurements with (1), using fixed nominal vehicle 3 parameters. Write $F^0(v,T) = \eta_T^0 T/R^0 - C_A^0 v^2 - C_r^0 m^0 g$, where superscript 0 denotes the nominal model. Then

$$\begin{aligned} r_p(k) &= \widetilde{p}_3(k) - \widetilde{p}_3(k-1) - \Delta t\, \widetilde{v}_3(k-1), \\ r_v(k) &= \widetilde{v}_3(k) - \widetilde{v}_3(k-1) \\ &\quad - \frac{\Delta t}{m^0}F^0\big(\widetilde{v}_3(k-1), \widetilde{T}_3(k-1)\big), \\ r_T(k) &= \widetilde{T}_3(k) - \widetilde{T}_3(k-1) \\ &\quad - \frac{\Delta t}{\tau^0}\big(u_3(k-1) - \widetilde{T}_3(k-1)\big). \end{aligned} \tag{4}$$

Four spacing/speed features are $\epsilon_{p,2}$, $-\epsilon_{v,2}$, $\epsilon_{p,1}$, and $-\epsilon_{v,1}$. Two torque-response features are $u_3 - \widetilde{T}_3$ and $T_3(k) - \widetilde{T}_3(k-1)$. Finally, each link contributes

$$c_{j3} = [\ell_{p,j3}, \ell_{v,j3}, a_{p,j3}, a_{v,j3}, d_{j3}, \iota_{j3}]^\mathsf{T}, \tag{5}$$

comprising field-specific update loss and information age, link delay, and identity inconsistency. Thus the group sizes

are $(8, 3, 4, 2, 12)$. No attack schedule or true plant state is supplied to the network.

The label order is FDI-position, FDI-velocity, FDI-torque, DoS-position, DoS-velocity, spoof-position, replay-position, and replay-velocity. A window $X_k = [\bar{z}(k-19), \ldots, \bar{z}(k)]^{\mathsf{T}}$ is assigned the eight-bit observable vector $y(k)$ at its final sample. Normal operation is $y = 0$. Attempted labels are retained separately; a remote attack is observable only when its modification or blocking action reaches the receive process. Concurrent remote mechanisms affecting the same physical field are allocated to different predecessor links. This constraint preserves a coherent supervised target without claiming identification of all $2^8$ possible sets. Observability here describes whether the simulated mechanism reaches the receiver, not a proof that its statistical signature is uniquely identifiable: benign loss can resemble DoS, and a local bias can resemble FDI. The classifier must resolve these ambiguities from history and complementary features. The labels identify mechanism and field, not attacker identity or the attacked predecessor link.

### *B. MOSAIC Algorithm*

Feature normalization uses only training episodes. For feature $r$, $\bar{z}_r = (z_r - \mu_r)/s_r$, where $\mu_r$ and $s_r$ are training-sample statistics; a nearly constant feature uses $s_r = 1$. The frozen transformation is used for validation, testing, and online inference. Let $X^{(g)}$ denote group $g$ of a normalized 20-sample window. Separate affine projections produce eight-dimensional embeddings:

$$\begin{aligned} U_t^{(g)} &= \tanh\Big(W_g X_t^{(g)} + b_g\Big), \\ h_{20} &= \mathrm{GRU}_{32}\Big([U_t^{(1)}; \cdots; U_t^{(5)}]_{t=1}^{20}\Big). \end{aligned} \tag{6}$$

The one-layer GRU receives 40 channels. For the concatenated embedding $u_t \in \mathbb{R}^{40}$, the implementation uses reset and update gates $r_t$ and $\zeta_t$, a candidate state $\nu_t$, and the recurrence

$$\begin{aligned} r_t &= \mathrm{sigmoid}(A_r u_t + B_r h_{t-1} + b_r), \\ \zeta_t &= \mathrm{sigmoid}(A_z u_t + B_z h_{t-1} + b_z), \\ \nu_t &= \tanh\big(A_n u_t + b_{in} + r_t \odot (B_n h_{t-1} + b_{hn})\big), \\ h_t &= (1 - \zeta_t) \odot \nu_t + \zeta_t \odot h_{t-1}. \end{aligned} \tag{7}$$

All gates and states have 32 entries, and $\odot$ denotes element-wise multiplication. The reset gate acts after the hidden affine map, as in the PyTorch GRU used in the experiments. Each window starts at $h_0 = 0$; hidden states are not transferred between different episodes or between overlapping windows. Consequently, training and online inference evaluate the same finite-history map rather than different recurrent-state conventions.

Layer normalization (LN) and a rectified linear unit (ReLU) yield $s = \mathrm{ReLU}(W_s\,\mathrm{LN}(h_{20}) + b_s) \in \mathbb{R}^{32}$. Each attack $i$ queries the window-averaged group evidence:

$$\begin{aligned} e_g &= 20^{-1} \sum_{t=1}^{20} U_t^{(g)}, \\ \alpha_{ig} &= \frac{\exp(q_i^{\mathsf{T}} e_g/\sqrt{8})}{\sum_{r=1}^{5} \exp(q_i^{\mathsf{T}} e_r/\sqrt{8})}, \\ c_i &= \sum_{g=1}^{5} \alpha_{ig} e_g, \qquad o_i = w_i^{\mathsf{T}}[s; c_i] + b_i. \end{aligned} \tag{8}$$

The queries select group evidence; they are not temporal self-attention or a causal explanation of the attacker. Eight sigmoid probabilities $\pi_i = \mathrm{sigmoid}(o_i)$ produce $\widehat{y}_i = \mathbb{1}\{\pi_i \geq \theta_i\}$. Shared recurrent dynamics couple the outputs, while the sigmoid heads allow multiple positive labels.

The architecture factorizes the output by attack mechanism rather than assigning a separate softmax class to every attack set. Thus an unseen triple can activate three previously trained heads. This construction reduces the output dimension, but does not guarantee combination transfer: concurrent feedback effects may change the embeddings and move several logits across their thresholds. The per-attack context is intended to make evidence selection explicit; the uniform-context and no-query variants provide direct tests of whether this additional structure is useful. No attention penalty, explicit label-correlation loss, or attack-priority postprocessing is imposed.

Algorithm 1 keeps three information boundaries explicit. Receive buffers determine observable labels before supervised learning; trajectory families are separated before normalization; and validation alone selects weights and thresholds. The positive-weighted binary cross-entropy (BCE) loss compensates for rare active labels without replacing the physical feasibility constraints. The Adam optimizer with decoupled weight decay (AdamW) updates the trainable maps. During inference, a first-in, first-out (FIFO) buffer retains only past and current measurements. Resetting this buffer at an episode boundary prevents unrelated histories from entering a window. The same frozen window map is applied to unseen combinations, so any transfer reflects learned evidence rather than additional fitting.

Training minimizes positive-weighted BCE:

$$\mathcal{L} = -\frac{1}{8B} \sum_{b=1}^{B} \sum_{i=1}^{8} \left[\omega_i y_{bi} \log \pi_{bi} + (1 - y_{bi}) \log(1 - \pi_{bi})\right], \tag{9}$$

where $\omega_i = \mathrm{clip}(N_i^- / \max(N_i^+, 1), 1, 30)$ is determined from training window labels. Algorithm 1 separates weight fitting, checkpoint selection, and threshold calibration. Thresholds maximize each validation-label F1 over $\{0.3, 0.5, 0.7\}$, with ties resolved toward $0.5$. Neither unseen combinations nor shifted-domain episodes participate in these steps. Online inference starts after 20 samples and advances once per new sample. Receiver features enter the FIFO in sampling order. The saved feature order, scaler, weights, label order, and thresholds reproduce the same decision rule after reloading. Empty output sets remain normal predictions.

MOSAIC has 8,888 trainable parameters: 272 in the group projections, 7,104 in the GRU, 64 in layer normalization, 1,056

**Algorithm 1** MOSAIC training and causal identification

**Require:** Nominal plant; feasible sets; disjoint trajectory families
**Ensure:** Frozen detector, scaler, thresholds; concurrent attack vector

**Causal data and target construction**
Allocate same-field remote mechanisms to distinct links.
**for** each simulated episode and sample $k$ **do**
  Apply local corruption; update field-wise receive buffers (3).
  Hold lost fields; advance age; constrain $s_k \le \max(0, k-1)$.
  Store attempted and receiver-observable labels separately.
  Compute $u_3$ via (2); form $z(k)$; advance the plant (1).
**end for**
Partition whole trajectory families, never overlapping windows.
Fit $(\mu, s)$ on training only; freeze the feature/label order.
Form 20-sample windows within episodes; target $y(k)$.
**Weight fitting and validation selection**
Initialize five projections, $\mathrm{GRU}_{32}$, queries, and eight heads.
Set $\omega_i = \mathrm{clip}(N_i^- / \max(N_i^+, 1), 1, 30)$ from training.
Initialize best validation score and non-improvement count.
**for** each training epoch **do**
  **for** each training minibatch **do**
    Set $h_0 = 0$; embed groups and evaluate (6)–(8).
    Minimize weighted BCE (9); clip gradient norm to 2.
    Apply AdamW; do not update the frozen scaler.
  **end for**
  Evaluate validation macro-F1 with all $\theta_i = 0.5$.
  **if** validation score improves **then**
    Save weights; reset non-improvement count.
  **else**
    Increment count; stop if four epochs do not improve.
  **end if**
**end for**
**Calibration and independent evaluation**
Reload best; maximize label F1 over $\theta_i \in \{0.3, 0.5, 0.7\}$.
Resolve ties toward 0.5; save scaler, orders, weights, and $\theta$.
Freeze all artifacts; score independent test protocols.
**Causal online identification**
Reload and check orders; clear FIFO at each episode boundary.
**for** each new receiver sample $k$ **do**
  Check finite $z(k) \in \mathbb{R}^{29}$; append to the 20-sample FIFO.
  **if** FIFO is full **then**
    Normalize the window; set $h_0 = 0$ and infer $\pi_i$.
    Return all $\widehat{y}_i = \mathbb{1}\{\pi_i \ge \theta_i\}$; an empty set is normal.
  **end if**
**end for**

in the shared affine layer, 64 in the queries, and 328 in the attack heads. The comparison Tiny-GRU directly receives 29 channels and uses the same 32-unit recurrent state and eight outputs, totaling 7,432 parameters. Both are small temporal classifiers; their resource difference is measured rather than assumed to favor MOSAIC.

## IV. Comparative Evaluations

### A. Concurrent Identification and Unseen Combinations

The evaluation uses verified synthetic closed-loop episodes. Vehicle 3 has nominal mass 1276 kg, torque lag 0.5828 s, drag coefficient 1.0352 kg/m, and tire radius 0.2638 m. Simulation settings are $d_0 = 20$ m, $C_r = 0.02$, $\eta_T = 0.95$, and torque/speed limits of $\pm 5000$ N·m and $[0, 30]$ m/s. Initial speed is uniform over $[16, 22]$ m/s. Each vehicle parameter varies independently within $\pm 2\%$. Gaussian measurement-noise standard deviations are $(0.05\,\mathrm{m}, 0.04\,\mathrm{m/s}, 12\,\mathrm{N\cdot m})$. Benign field loss is 1.5%, and delay is uniform over $[0.005, 0.06]$ s, converted to a causal lag of at least one sample. Identity inconsistency has probability 0.002 normally and 0.60 during spoofing, avoiding a perfect label indicator.

Each 20-s episode has 201 samples at $\Delta t = 0.1$ s and yields 182 windows. Attack onset is sampled from indices 50–130 and duration from 36–84 samples, clipped at episode end. Nominal signed FDI amplitudes have magnitudes $[2.5, 8]$ m, $[1, 4]$ m/s, and $[350, 1100]$ N·m; spoof amplitudes are $[5, 14]$ m. These amplitudes multiply step, ramp, sinusoidal, drifting, or random-walk profiles; the random-walk multiplier is clipped to $[0.15, 1.35]$. Replay adds 6–17 lag samples. Pair onsets are staggered by three samples and overlap for at least 15 samples.

P0 trains and tests normal/single-attack episodes. P1 includes normal operation, eight singles, and all 28 feasible pairs: 112/55/56 training/validation/test episodes yield 20,384/10,010/10,192 windows. P2 freezes the P1 models and evaluates three fresh episodes for each of four held-out sets:

$$\begin{gathered}\{\text{FDI-p}, \text{DoS-v}, \text{Replay-p}\},\\ \{\text{FDI-v}, \text{DoS-p}, \text{Replay-v}\},\\ \{\text{FDI-T}, \text{Spoof-p}, \text{DoS-v}\},\\ \{\text{FDI-p}, \text{FDI-T}, \text{Spoof-p}, \text{Replay-v}\}.\end{gathered} \tag{10}$$

These 12 episodes supply 2,184 evaluated windows. P3 uses 56 independent shifted-domain episodes. All models use eight epochs, batch size 256, AdamW learning rate $10^{-3}$, weight decay $10^{-4}$, and gradient-norm limit 2.0. The P1 training/data seed is 11; P0 is trained independently. This is a single-seed study, not a multi-seed performance estimate.

Table I reports macro-/micro-F1 and complete eight-bit subset accuracy. On P1, MOSAIC improves macro-F1 by 3.62 percentage points and reduces Hamming loss from 2.99% to 2.52%. Its normal-window false-positive rate is 3.82%, versus 4.97% for Tiny-GRU. Fig. 2 reveals uneven gains: velocity FDI improves from 60.06% to 75.68% F1, whereas position replay decreases from 82.51% to 72.81%. Position FDI remains weakest at 40.72% F1; the aggregate gain does not benefit every attack.

The added label accuracy equals one minus Hamming loss; subset accuracy requires every label in a window to be correct. Micro-precision and micro-recall pool positive decisions across attack labels before forming their ratios. On P2, MOSAIC raises recall from 57.23% to 71.24%, with precision changing from 82.40% to 81.20%, indicating fewer missed mechanisms rather than indiscriminate gains. On P3, recall reaches 85.62%, but precision is only 48.27%, exposing substantial false alarms. High label accuracy partly reflects inactive labels and must not substitute for complete-set recovery. The additional measures therefore distinguish missed mechanisms, false alarms, and reliable joint identification under each protocol.

P2 macro-F1 improves by 9.91 points. However, 1,351 of its 2,184 windows have no active label, so Fig. 3 isolates concurrent intervals. MOSAIC obtains 80.13% micro-F1 and 29.90% subset accuracy on triple-label windows; four-label windows yield 70.58% and 0%. Fig. 4 shows the predeter-

TABLE I
PROTOCOL SCORES FROM INDEPENDENT TEST EPISODES (%)

| Protocol | Model | Macro-F1 | Micro-F1 | Subset accuracy | Label accuracy | Micro-precision | Micro-recall |
|---|---|---|---|---|---|---|---|
| P0 | Tiny-GRU | 79.45 | 77.14 | 85.71 | 98.10 | 69.22 | 87.09 |
| | MOSAIC | 80.45 | 77.62 | 86.97 | 98.18 | 71.07 | 85.50 |
| P1 | Tiny-GRU | 75.29 | 74.66 | 81.43 | 97.01 | 72.01 | 77.53 |
| | MOSAIC | 78.91 | 78.64 | 84.07 | 97.48 | 75.73 | 81.78 |
| P2 | Tiny-GRU | 60.89 | 67.55 | 64.47 | 92.25 | 82.40 | 57.23 |
| | MOSAIC | 70.81 | 75.90 | 69.41 | 93.62 | 81.20 | 71.24 |
| P3 | Tiny-GRU | 48.34 | 38.23 | 34.78 | 87.08 | 27.00 | 65.45 |
| | MOSAIC | 65.69 | 61.74 | 61.45 | 93.52 | 48.27 | 85.62 |

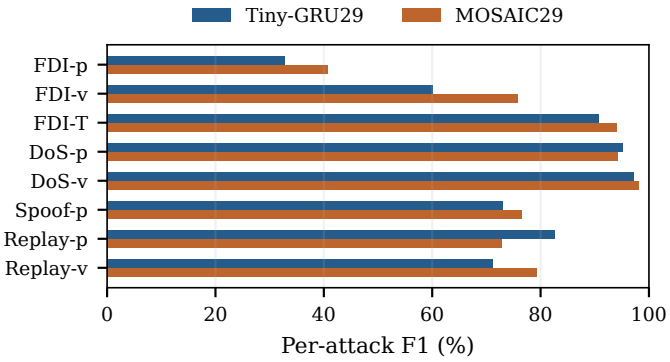


Fig. 2. P1 per-attack F1. The grouped representation helps several labels, but position replay is weaker than the direct-input baseline.

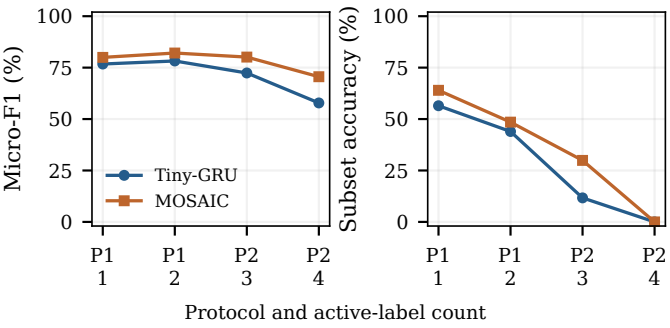


Fig. 3. Scores conditioned on active-label count. P1 contributes 1,560/1,536 single/pair windows; P2 contributes 505/180 triple/quadruple windows. The protocols are evaluated separately.

mined repetition-0 quadruple: individual mechanisms are often detected, but the complete set is not recovered. Under the first-positive event rule, position replay misses one of ten P1 events; mean delay is 0.44 s among nine detected events. These results distinguish sustained labeling, joint recovery, and event detection.

A paired bootstrap resamples the same 56 P1 episode blocks 2,000 times, retaining overlapping windows within each block. The 95% macro-F1 intervals are $[68.62, 80.01]\%$ for Tiny-GRU and $[73.37, 82.67]\%$ for MOSAIC. The paired difference interval, $[-1.16, 8.50]$ percentage points, includes zero. The P1 improvement is therefore descriptive evidence from one training realization, without established statistical superiority.

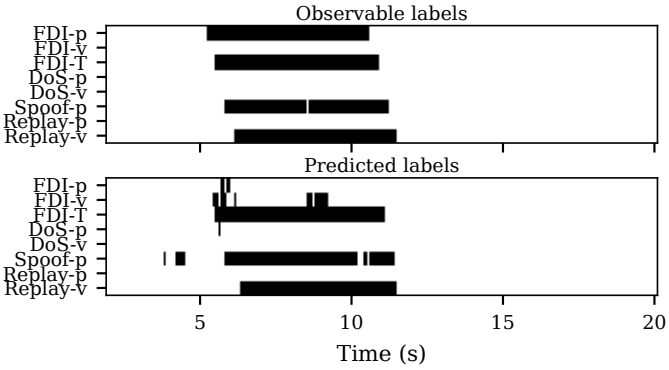


Fig. 4. Observable and inferred labels for a held-out FDI-position/FDI-torque/spoof-position/replay-velocity quadruple. Black denotes an active label at the window endpoint.

TABLE II
P1 ABLATION AND ARCHITECTURE COST

| Variant | Structure | Parameters | Macro-F1 (%) |
|---|---|---|---|
| A0 | Tiny-GRU29 | 7,432 | 75.29 |
| A1 | Tiny-GRU25 | 7,048 | 61.07 |
| A2 | Projected GRU | 8,760 | 79.27 |
| A3 | Full MOSAIC | 8,888 | 78.91 |
| A4 | No V2V group | 8,016 | 48.59 |
| A5 | Uniform aggregation | 8,824 | 81.10 |

### *B. Ablation, Domain Shift, and Computational Cost*

Table II reports independently retrained variants on the same P1 episode split. A1 removes velocity-loss and velocity-age channels from both V2V links, yielding a 25-dimensional eight-output baseline; it is not the earlier four-head anomaly task. A2 retains all five projections and the GRU but uses a shared eight-output layer without queries or group context. A4 removes the entire communication group. A5 retains group context with uniform weights $\alpha_{ig} = 1/5$.

Removing communication evidence costs 30.31 points relative to A3. The smaller A1 input also loses 17.83 points, indicating that field-specific metadata helps distinguish loss and staleness in this simulator. By contrast, A2 exceeds A3 by 0.36 points, and uniform aggregation exceeds it by 2.20 points. The learned-query component therefore has no demonstrated benefit in this run. Fig. 5 shows mean group weights on label-positive windows; their proximity to uniform weights is consistent with this result, although the weights alone do not establish feature causality. Longer and repeated training is

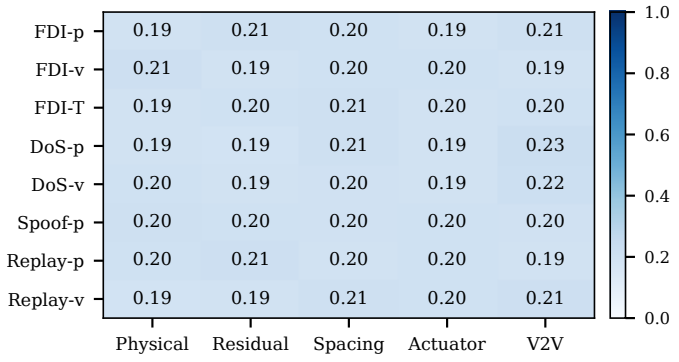


Fig. 5. Mean learned group weights on positive P1 windows. The color scale is fixed to $[0, 1]$; small differences are not magnified.

needed to determine whether the query mechanism improves combination transfer or merely adds estimation variance.

P3 combines an unseen emergency leader profile with $\pm 10\%$ parameter perturbations, 1.5 times the baseline sensor-noise standard deviations, 6% benign field loss, and delay in $[0.01, 0.12]$ s. Additive FDI/spoofing uses held-out chirp and burst waveforms; DoS and replay retain their loss/staleness mechanics. The nominal residual model, scaler, checkpoint, and thresholds remain fixed. MOSAIC retains 65.69% macro-F1, exceeding the baseline by 17.36 points, but loses 13.21 points relative to P1. Its normal-window false-positive rate increases to 29.89%; the baseline reaches 63.72%. Higher label recall under this shift must therefore be interpreted together with false alarms. The study supports partial transfer to a harder synthetic domain, not dependable deployment under arbitrary operating uncertainty.

Experiments use Python 3.12.11 and PyTorch 2.7.1 with Compute Unified Device Architecture (CUDA) 11.8 on an NVIDIA GeForce RTX 3060 Laptop graphics processing unit (GPU). The implementation's approximate operation count treats each multiply-add as two floating-point operations (FLOPs) and omits nonlinearities, context aggregation, and runtime overhead. Tiny-GRU and MOSAIC require 236,800 and 289,088 counted FLOPs/window; 32-bit floating-point (FP32) weights occupy 29.03 and 34.72 kibibytes (KiB). After warm-up, 100 batch-one central processing unit (CPU) measurements yield median/95th-percentile (p95) latency of 0.219/0.255 ms and 0.561/0.998 ms (Fig. 6). Synchronized GPU medians are 0.195 and 0.677 ms. MOSAIC is compact, but neither faster nor smaller than Tiny-GRU. Timings cover host classifier inference, not V2V acquisition or embedded execution.

Nineteen automated checks cover causality, observable labels, normalization, split isolation, and model dimensions; checkpoint reloading reproduces P1. The full episode budget and independent training seeds remain unevaluated. Field-addressable packets are assumed, and short synthetic episodes do not establish long-term false-alarm behavior. External traces, authenticated V2V metadata, alternative controllers, and hardware measurements are needed to assess practical reliability.

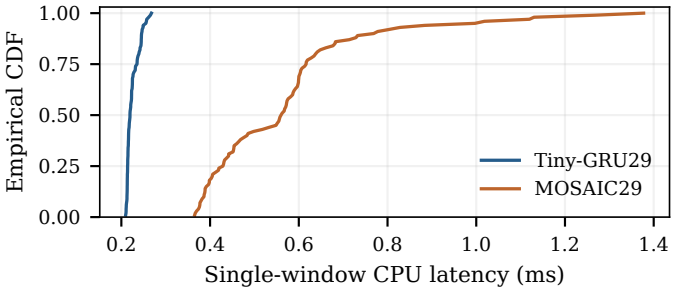


Fig. 6. Empirical cumulative distribution function (CDF) of central processing unit (CPU) inference latency for one 20-sample window, after warm-up, with 100 measurements per model.

## V. Conclusion

MOSAIC combines receiver-observable multi-label targets with grouped temporal evidence for concurrent CAEV cyber-attacks. The compact model improves seen-pair and held-out-combination F1 over a direct-input Tiny-GRU in the evaluated simulation. Complete label-set recovery remains difficult, particularly for four concurrent attacks, and shifted domains produce substantial false alarms. Communication evidence contributes strongly; learned attention does not outperform uniform aggregation in this study. These findings define a reproducible starting point for further combination-transfer analysis, repeated training, and independent vehicle-data validation.

## References

[1] Z. Xu, Y. Wang, Z. Wu, W. X. Zheng, and C. Hu, "Federated learning-based offset-free distributed control of nonlinear networked systems with application to IIoT," *IEEE Trans. Netw. Sci. Eng.*, vol. 12, no. 3, pp. 1859–1871, 2025, doi: 10.1109/TNSE.2025.3540643.

[2] H. Wang, H. Luo, L. Ren, M. Huo, Y. Jiang, and O. Kaynak, "Data-driven design of distributed monitoring and optimization system for manufacturing systems," *IEEE Trans. Ind. Informat.*, vol. 20, no. 7, pp. 9455–9464, 2024, doi: 10.1109/TII.2024.3383491.

[3] Z. Xu, X. Zhang, L. Zhang, Y. Xie, Z. Chen, and D. Wang, "Fault diagnosis and initial alignment of redundant SINS under large misalignment angle," *IEEE Trans. Ind. Informat.*, vol. 22, no. 7, pp. 5678–5687, 2026, doi: 10.1109/TII.2026.3670431.

[4] S. Lu, Z. Gao, Q. Xu, C. Jiang, A. Zhang, and X. Wang, "Class-imbalance privacy-preserving federated learning for decentralized fault diagnosis with biometric authentication," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 9101–9111, 2022, doi: 10.1109/TII.2022.3190034.

[5] F. van Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1264–1276, 2020, doi: 10.1109/TITS.2019.2906038.

[6] Z. Zhang, Y. Yao, W. Hutabarat, M. Farnsworth, D. Tiwari, and A. Tiwari, "Time series anomaly detection in vehicle sensors using self-attention mechanisms," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 11, pp. 15964–15976, 2024, doi: 10.1109/TITS.2024.3415435.

[7] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, 2014, doi: 10.1109/TKDE.2013.39.

[8] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, 2013, doi: 10.1109/TAC.2013.2266831.

[9] Q. Chen, D. J. Sun, J. Xia, X. Liu, and C. An, "String stable and robust optimal longitudinal control for vehicle platoons: a finite time disturbance observer-based tube MPC," *Nonlinear Dyn.*, vol. 113, pp. 16593–16610, 2025, doi: 10.1007/s11071-025-10964-0.