Federated Analytics Hybrid Architecture for Connected Truck Data Analysis

Aslihan Reyhanoglu[1], Feyzi Ege Kumec[1], Abdulkadir Bilge[1], Hira Bakirhan[2], Bugra Turan[3], and Sinem Coleri[3]

[1]Koc University Ford Otosan Automotive Technologies Laboratory (KUFOTAL), Sariyer, Istanbul, Turkey
[2]Ford Otosan, Sancaktepe, Istanbul, Turkey
[3]Department of Electrical and Electronics Engineering, Koc University, Sariyer, Istanbul, Turkey

E-mail: areyhanoglu@ku.edu.tr

2026

About the Presenter

- Aslıhan Reyhanoglu — AI R&D Engineer (Koc University), Istanbul
- Focus area: applied ML/AI, telemetry-driven analytics, predictive modeling, anomaly monitoring, communication networks, vehicular communications and connected vehicle systems.
- Education: MSc in Radio and Mobile Communication Systems; BSc in Electronic and Communication Engineering
- Research areas and publications:
    - C-V2X Mode 4 reliability analysis (IEEE VTC-Fall 2022)
    - NR-V2X QoS prediction (IEEE VNC 2023)
    - Federated learning for pedestrian detection (IEEE BlackSeaCom 2023)
    - Network-aware bitrate prediction for remote driving (IEEE BlackSeaCom 2025)

Outline

Introduction

- Connected vehicles offer opportunities to optimize fleet operations by leveraging data from their on-board sensors.
- For remote vehicle health monitoring, the 5G Automotive Association (5GAA) defines a latency of less than 30 seconds as acceptable.
- Federated Analytics (FA) enables decentralized data analysis to address latency, scalability, and privacy challenges in connected vehicle systems by combining localized edge computing with cloud-based aggregation.
- In an FA approach, raw data remains on the vehicles and only compact analytical outputs are shared with the cloud.

Problem Statement

- Complex data and heavy transmission demands can overwhelm centralized monitoring systems.
- This data overload leads to latency, network congestion, and costly cloud storage.
- Existing studies have not fully addressed real-time truck health monitoring using processed anomalies without requiring heavy on-board model retraining.

Contributions

- A practically deployable edge–cloud FA architecture for connected truck health monitoring.
- On-board processing on NVIDIA Jetson Orin: high-rate telemetry $\rightarrow$ compact warning/event summaries.
- Cloud-side Temporal–Spatial Aggregation (TSA) to produce near real-time hotspot maps using 15-minute windows over geo-fenced zones.
- Experimental validation: near real Ford F-Max test truck data + synthetic controlled tests, achieving $\sim$1s E2E latency in evaluated settings.

System Model: Federated Analytics Architecture

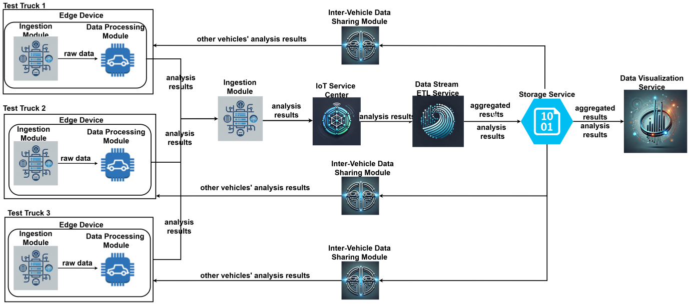

Figure: Proposed Federated Analytics (FA) Architecture.

- Edge modules run on Jetson Orin; cloud handles long-term storage + fleet-level visualization/analytics.
- Overall flow: raw telemetry → on-board processing → warning/event summaries → IoT Service Center → ETL → storage → dashboards and hotspot visualization.

Modules

Module	**Role**
Ingestion Module	Collects and streams raw telemetry (ROS)
Data Processing Module	On-board analytics; outputs warnings/events
IoT Service Center	Receives and routes edge outputs
ETL Service	Parses, enriches, and transforms incoming messages for storage and analytics
Storage Service	Stores outputs and aggregations
Inter-Vehicle Data Sharing	Allows trucks to exchange processed data via cloud
Visualization Service	Dashboards and hotspot views

Baseline: Raw Data Transfer Architecture

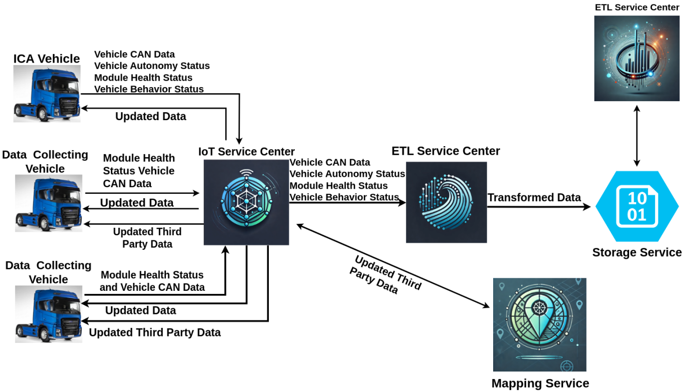

Figure: Baseline architecture: raw telemetry sent to cloud (no edge analytics).

- Baseline transmits raw streams end-to-end; used to benchmark latency vs. FA.

On-board Data Processor

- Runs independently on each vehicle to detect abnormal behavior and system anomalies in real time.
- Deterministic, configurable **rule-based threshold checks** (domain heuristics).

Event Type	Detection Rule (threshold & duration)
Aggressive acceleration	Throttle $> 90\%$ for > 15 s
Harsh braking	Brake $> 80\%$ more than 3 times within 5 min
Erratic steering	Steering angle $> 30^\circ$ at least 3 times within 10 s
Lane-departure risks	Yaw error > 0.05 rad **or** lateral deviation > 0.5 m
Module faults	Repeated sensor faults (camera/radar/traffic predictor) over consecutive readings
Abrupt control transitions	Mode changes within 2 s

- Outputs compact Warning JSON with vehicle ID, warning type, and event hits (timestamps + GPS).

Off-board Temporal–Spatial Aggregation for Hotspots

- Aggregates processed warnings in real time at the cloud backend.
- Maps events to geo-fenced zones using GPS coordinates (routes/test areas).
- Uses a 15-minute sliding window updated every 1 minute.
- Computes hotspot metrics per zone+window:
  - Total number of threshold-exceeding events (e.g., high braking events)
  - Number of unique vehicles involved (active vehicles)
- Output schema (example): `z, ws, we, hb, av` enabling near real-time fleet visibility.

Experimental Setup

- Goal: benchmark responsiveness and system impact via latency metrics.
- Metrics:
  - **On-board** $\rightarrow$ **Off-board** latency (vehicle to backend transmission)
  - **IoT Service Center** $\rightarrow$ **ETL** latency (cloud pipeline processing)
  - **Average E2E** latency (Vehicle $\rightarrow$ Cloud $\rightarrow$ Vehicle)
- Two setups:
  - **Real-world comparison** of baseline vs. FA using instrumented Ford F-Max test trucks (unidirectional; repeated 5 times).
  - **Synthetic bidirectional tests** under varying data frequencies; telemetry messages $\sim$3311 bytes; FA analysis results up to 134 bytes.

Latency: Baseline vs. FA (Real Truck Data)

- Compared average end-to-end latency between raw data transfer and FA architecture.
- Key observation: FA adds overhead (processing + storage retrieval), but greatly reduces data volume by sending event summaries instead of raw streams.

Topic	**Raw (s)**	**FA (s)**
Module status (20Hz)	0.55	1.03
HP vehicle status (100Hz)	0.62	1.11
Combined streams	0.746	1.12

Table: Average E2E latency comparison (paper Table II).

Data Volume Reduction

- Raw message sizes reported in the real-world baseline comparison:
  - Module status: ∼180 bytes
  - HP vehicle status: ∼1850 bytes
- In this setup, FA transmits compact analytics outputs (event summaries) as JSON messages of up to ∼200 bytes.
- *Note: In the separate synthetic bidirectional setup, raw telemetry messages are ∼3,311 bytes, while retrieved FA analysis results reach up to 134 bytes.*

Frequency Trade-off (Synthetic Bidirectional Tests)

- Latency evaluated under different data frequencies (0.5–20 Hz) in bidirectional operation.
- Optimal region observed around **5–10 Hz** (stable E2E latency); at 20 Hz, latency increases due to load/queuing.

Data Freq.	V1 → V2			V2 → V1		
	On→Off	**IoT→ETL**	**Avg. E2E**	**On→Off**	**IoT→ETL**	**Avg. E2E**
0.5 Hz	0.083s	0.107s	1.18s	0.009s	0.109s	1.07s
1 Hz	0.018s	0.108s	1.05s	0.009s	0.104s	0.87s
5 Hz	0.008s	0.110s	0.76s	0.008s	0.104s	0.76s
10 Hz	0.009s	0.105s	0.80s	0.008s	0.104s	0.79s
20 Hz	0.009s	0.109s	0.86s	0.128s	0.114s	1.10s

Table: Latency analysis under different data frequencies (paper Table III).

Conclusions

- Presented an FA-assisted edge–cloud platform for real-time monitoring in connected truck fleets.
- Edge analytics on Jetson Orin produces explainable warnings and reduces raw data transfer by sending compact summaries.
- Cloud TSA provides time- and location-correlated hotspot insights for fleet-level operational awareness.
- Experiments demonstrate near $\sim$1s E2E latency in evaluated configurations and highlight frequency vs. compute/network trade-offs.

Future Work

- Latency optimization via tighter edge–cloud synchronization and (potentially) 5G integration.
- Lightweight unsupervised learning to dynamically adjust thresholds based on driver profiles and environmental context.
- Further robustness improvements for transient noise and high-frequency workloads.

Questions

Thanks for your attention! Any questions?

Aslihan Reyhanoğlu
areyhanoglu@ku.edu.tr