World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(Review Article)

Vehicular data management at scale: Architectural frameworks for cars as mobile data centers

Mohammed-Javed Padinhakara *

Independent Researcher, USA.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(03), 752-758

Publication history: Received on 27 April 2025; revised on 01 June 2025; accepted on 04 June 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.3.0940

Abstract

The emerging paradigm of modern vehicles as sophisticated mobile data centers generates unprecedented volumes of telemetry, sensor, and interaction data that require novel management approaches. The architectural framework addresses dual requirements of edge processing for latency-sensitive applications and cloud infrastructure for deeper analytics and model development. Vehicle-to-everything communication protocols integrate with software-defined networks and distributed ledger technologies to ensure secure, efficient data exchange across the ecosystem. Technical challenges including bandwidth constraints, data redundancy, and privacy regulations are primary motivators for solutions based on federated learning, optimized compression algorithms, and context-aware processing. Resilient vehicular data management necessitates a multi-layered approach balancing computational requirements across the edge-cloud continuum while maintaining robust security postures. These foundations enable scaling next-generation intelligent transportation systems were vehicles function as key nodes in broader smart city infrastructures.

Keywords: Vehicular Data Management; Edge Computing; V2X Communication; Federated Learning; Data Privacy

1. Introduction

The modern automobile has undergone a profound transformation from a primarily mechanical system to an increasingly sophisticated mobile data center on wheels. This evolution represents a paradigm shift in how vehicles are designed, operated, and integrated into broader transportation networks [1]. Contemporary vehicles now house numerous data-generating components, including advanced driver-assistance systems (ADAS), connected infotainment platforms, and emerging autonomous driving capabilities, collectively transforming cars into complex computing environments that continuously produce, process, and transmit data.

1.1. Evolution of Modern Vehicles into Mobile Data Centers

The proliferation of vehicular data systems has created an unprecedented expansion in the volume, variety, and velocity of automotive data. This transformation has accelerated with the emergence of the Internet of Vehicles (IoV), wherein connected cars serve as critical nodes in a larger intelligent transportation ecosystem [1]. Each vehicle now incorporates numerous sensors ranging from cameras and radar to ultrasonic and LiDAR systems, all generating continuous streams of environmental and operational data.

1.2. Proliferation of Data-Generating Systems in Automobiles

Modern vehicles integrate an expanding array of data-generating systems. ADAS features such as adaptive cruise control, lane keeping assistance, and emergency braking systems continuously capture and process sensor data to ensure safe vehicle operation. Simultaneously, connected infotainment systems provide multimedia experiences while

* Corresponding author: Mohammed-Javed Padinhakara

collecting user interaction data and maintaining network connectivity. As vehicles progress toward higher levels of autonomy, the computational demands and data throughput requirements increase exponentially [2].

1.3. Growing Challenges of Vehicular Data Management at Scale

With this exponential growth in vehicular data generation comes a corresponding set of challenges in data management at scale. The automotive industry now faces complex issues related to on-board processing capacity, network bandwidth constraints, storage optimization, and computational resource allocation. Even fundamental connectivity architectures like PCI Express technology have had to evolve to accommodate the increasing data throughput demands from infotainment systems to ADAS applications [2]. These challenges are further complicated by requirements for real-time processing, security, privacy compliance, and system resilience.

1.4. Research Objectives and Significance

This research examines the architectural frameworks necessary to address these vehicular data management challenges at scale. We investigate optimized approaches for processing vehicular data across the edge-cloud continuum, with particular emphasis on balancing computational requirements between in-vehicle processing and cloud infrastructure. The significance of this research lies in establishing foundational architectural principles that will enable the next generation of intelligent vehicles and transportation systems to effectively manage the growing data deluge while maintaining performance, security, and privacy requirements.

2. Vehicular Data Ecosystem Architecture

The architecture supporting vehicular data management represents a complex ecosystem spanning from in-vehicle systems to cloud infrastructure. This ecosystem must efficiently handle diverse data types while maintaining performance guarantees for time-critical applications and enabling deeper analytics for non-time-sensitive functions. The integration of edge, fog, and cloud computing paradigms creates a multi-tiered architecture that forms the backbone of modern vehicular data management systems.

2.1. Edge Computing Paradigms for In-Vehicle Processing

Edge computing has emerged as a pivotal approach for handling real-time vehicular data processing directly within the vehicle. This paradigm shifts computational resources closer to data sources, reducing latency for safety-critical functions while minimizing bandwidth requirements for external communication. As outlined by Herbert Raj, et al., device-based edge architectures enable vehicles to process sensor data locally, making immediate decisions without relying on external connectivity [3]. These edge systems utilize in-vehicle compute platforms with specialized hardware accelerators for machine learning inference, allowing ADAS applications to function reliably even in environments with intermittent network connectivity.

Table 1 Vehicular Computing Paradigms Comparison [3, 4]

Computing Paradigm	Processing Location	Latency	Computational Capacity	Typical Applications
Edge Computing	In-vehicle	Minimal	Limited	ADAS, emergency braking, collision avoidance
Fog Computing	Roadside infrastructure	Low	Moderate	Traffic coordination, local hazard warnings
Cloud Computing	Data centers	Higher	Extensive	Fleet analytics, HD map updates, ML model training

2.2. Cloud Infrastructure Integration for Analytics and Model Refinement

While edge computing addresses immediate processing needs, cloud infrastructure provides the computational capacity required for comprehensive data analytics and machine learning model refinement. The vehicle-to-cloud interface facilitates the transmission of aggregated telemetry data, driving patterns, and system performance metrics to centralized repositories. Sachin B. Chougule notes that cloud platforms enable automotive manufacturers to analyze fleet-wide patterns, refine AI models, and deploy optimized parameters back to vehicles through over-the-air updates

[4]. This bidirectional flow creates a continuous improvement cycle where vehicle performance benefits from collective learning across the entire fleet.

2.3. Vehicle-to-Everything (V2X) Communication Protocols

The functionality of connected vehicles extends beyond internal data processing to encompass interactions with surrounding infrastructure, other vehicles, and various network services through V2X communication protocols. These standardized communication frameworks enable vehicles to exchange critical safety information, traffic conditions, and environmental data with neighboring entities. The V2X ecosystem encompasses several communication modalities, including vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-pedestrian (V2P) interactions [3]. These protocols must balance data throughput requirements with stringent latency constraints while maintaining security against potential intrusions.

2.4. Software-Defined Networks (SDN) in Automotive Applications

The dynamic nature of vehicular networks requires flexible connectivity management, which has led to the adoption of software-defined networking approaches in automotive applications. SDN architectures separate the network control plane from the data forwarding plane, allowing for programmatic network management that can adapt to changing conditions. This flexibility becomes particularly valuable in handling the fluctuating bandwidth demands and connection transitions that characterize vehicular mobility. Herbert Raj, et al. highlight that SDN implementations enable prioritization of critical data streams while efficiently managing non-time-sensitive transmissions across available network interfaces [3].

2.5. Distributed Ledger Technologies for Secure Data Exchange

The secure exchange of vehicular data has become increasingly important as connected vehicles participate in broader transportation ecosystems. Distributed ledger technologies provide mechanisms for establishing trusted data provenance and secure multi-party transactions without relying on centralized authorities. Chougule examines how blockchain and related technologies can create tamper-resistant records of vehicle-generated data, enabling secure sharing among stakeholders while maintaining driver privacy [4]. These systems play a crucial role in supporting emerging mobility models where vehicles may participate in data marketplaces or contribute to collaborative applications like high-definition mapping.

3. Data Processing Challenges and Solutions

The exponential growth in vehicular data generation introduces significant processing challenges that require innovative solutions. As vehicles evolve into sophisticated mobile data centers, they must efficiently handle diverse data types while operating under mobile environment constraints. This section examines the primary challenges in vehicular data processing and explores strategic approaches to address these limitations while maintaining system performance.

3.1. Bandwidth Limitations in Mobile Environments

Vehicular networks operate within inherently constrained and fluctuating bandwidth environments. The mobile nature of vehicles means they regularly transition between different connectivity zones, from high-bandwidth urban areas to limited-connectivity rural regions. According to Andersen, G., & MoldStud Research Team, these bandwidth variations create significant challenges for maintaining consistent data transmission between vehicles and cloud infrastructure [5]. The cellular network coverage gaps, congestion during peak traffic periods, overhead during switchover between network types (5G/4G/3G), and interference patterns in urban environments further compound these limitations. Successful vehicular data architectures must implement adaptive transmission strategies that prioritize critical data streams during bandwidth constraints while deferring non-essential transmissions to more favorable connectivity conditions.

3.2. Data Redundancy Mitigation Strategies

Vehicular systems generate substantial redundant data across multiple sensors and subsystems. Yaping Zhang identifies this redundancy as a key inefficiency in automotive data flows that exacerbates bandwidth and storage constraints [6]. Redundancy manifests through overlapping sensor coverage, duplicate message transmissions, and repetitive telemetry reporting during stable operating conditions. Effective mitigation strategies include implementing intelligent data filtering at the source, context-aware sampling rates that adjust based on environmental conditions, and differential updates that transmit only changed values rather than complete state information. These approaches substantially reduce unnecessary data transmission without compromising system functionality or analytical capabilities.

3.3. Storage Optimization for Heterogeneous Vehicular Data

The diverse nature of vehicular data presents unique storage optimization challenges across the edge-cloud continuum. Each data type—from structured telemetry to unstructured sensor feeds—exhibits different access patterns, retention requirements, and analytical value. Zhang examines how tiered storage architectures can align storage medium characteristics with data attributes to optimize both cost and performance [6]. Time-sensitive operational data may be maintained in high-performance in-vehicle storage with limited retention periods, while aggregated historical data undergoes transformation before transfer to cloud repositories optimized for analytical workloads. Effective storage strategies must also address metadata management to maintain contextual information and ensure data usability across diverse applications.

Table 2 Vehicular Data Categories [5, 6]

Data Category	Examples	Temporal Relevance	Storage Location	Processing Priority
Safety-critical	Collision detection, brake system telemetry	Immediate	Edge (in-vehicle)	Highest
Operational	Engine performance, fuel efficiency	Short-term	Edge/Fog	High
Environmental	Road conditions, weather data	Medium-term	Fog/Cloud	Medium
User Experience	Navigation history, media preferences	Long-term	Cloud	Low

3.4. Real-Time vs. Batch Processing Considerations

The vehicular data ecosystem encompasses processing requirements that span from ultra-low-latency safety applications to long-term fleet analysis. Balancing these diverse processing modalities requires careful workload segregation and resource allocation. Andersen, G., & MoldStud Research Team highlight the critical distinction between stream processing for immediate decision-making and batch processing for deeper analytical insights [5]. Real-time processing typically occurs at the edge, leveraging specialized hardware accelerators to minimize latency for safety-critical functions. In contrast, batch processing operations can be scheduled during optimal connectivity periods or offloaded to cloud infrastructure where computational resources are more abundant. This bifurcated approach ensures critical applications maintain performance guarantees while still enabling comprehensive data analysis.

3.5. Compression Algorithms for Efficient Data Transmission

Compression techniques play a vital role in mitigating bandwidth constraints while maintaining data integrity across vehicular networks. According to Zhang, the selection of appropriate compression algorithms must consider not only compression ratios but also computational overhead, latency impact, and information preservation requirements [6]. Lossless compression approaches preserve complete data fidelity while achieving moderate space reduction for structured telemetry and diagnostic data. For sensor feeds where perfect reconstruction is less critical, lossy compression can achieve substantially higher compression ratios while preserving essential features. Adaptive compression frameworks that dynamically select algorithms based on data characteristics, network conditions, and application requirements provide the flexibility needed in dynamic vehicular environments.

4. Security and Privacy Frameworks

As vehicles evolve into sophisticated data platforms, robust security and privacy frameworks become essential components of vehicular data management. Modern vehicles must not only protect against malicious intrusions but also ensure compliance with increasingly stringent privacy regulations across global jurisdictions. This section examines the multifaceted approaches required to maintain data security and privacy throughout the vehicular data lifecycle.

4.1. Regulatory Compliance Requirements

The global landscape of data protection regulations presents significant compliance challenges for automotive systems that operate across jurisdictional boundaries. According to Compliance Hub, regulations such as the General Data

Protection Regulation (GDPR) in Europe, the California Consumer Privacy Act (CCPA) in the United States, and the Lei Geral de Proteção de Dados (LGPD) in Brazil impose specific requirements on vehicular data collection, storage, and processing [7]. These regulations establish fundamental principles including purpose limitation, data minimization, storage constraints, and individual rights to access, correct, and delete personal information. Vehicular data architectures must implement privacy-by-design approaches that embed compliance mechanisms directly into system operations rather than treating privacy as an afterthought.

4.2. Data Anonymization and Privacy-Preserving Techniques

Protecting user privacy while maintaining analytical utility represents a core challenge in vehicular data management. Ulf Mattsson, MSE highlights that effective privacy protection requires more than simple data masking, necessitating sophisticated techniques that preserve statistical properties while preventing re-identification [8]. Applied approaches include differential privacy that introduces calibrated noise to dataset queries, k-anonymity that ensures individuals cannot be distinguished within similarity groups, and homomorphic encryption that enables computation on encrypted data without decryption. These methodologies allow automakers to derive valuable insights from aggregated fleet data without compromising individual privacy, addressing both regulatory requirements and ethical considerations around personal data usage.

4.3. Authentication and Authorization Models for Vehicular Data

Securing access to vehicular data systems requires robust authentication and authorization frameworks that establish trusted identities and enforce appropriate access controls. As detailed by Mattsson, these systems must accommodate diverse access patterns across the vehicle ecosystem, from driver interactions with infotainment systems to maintenance access by service technicians [8]. Modern approaches leverage multi-factor authentication, context-aware access policies, and attribute-based access controls to create granular security models. The integrity of these systems depends on secure credential management, including protected storage for cryptographic materials and certificate-based identity verification for both human and machine entities within the vehicular network.

4.4. Threat Detection and Mitigation Strategies

The expanding attack surface of connected vehicles necessitates comprehensive threat detection and mitigation capabilities. Compliance Hub emphasizes that effective security requires a defense-in-depth approach combining preventive controls with active monitoring and incident response mechanisms [7]. Intrusion detection systems analyze network traffic patterns and system behaviors to identify potential security breaches, while anomaly detection algorithms flag unusual data access patterns that may indicate compromise. Segmentation strategies isolate critical vehicle systems from potentially vulnerable components, limiting attack propagation paths. Additionally, secure boot processes and code signing requirements ensure that only authorized software executes within vehicle environments, protecting against malware introduction.

4.5. Secure Over-the-Air (OTA) Update Mechanisms

The ability to remotely update vehicle software has transformed maintenance models while introducing new security considerations. According to Mattsson, secure OTA update systems must maintain update integrity throughout the distribution pipeline while preventing unauthorized modifications [8]. Cryptographic signing ensures that only authenticated updates from trusted sources are applied to vehicle systems. Delta updates that transmit only changed code components reduce bandwidth requirements while minimizing attack surfaces. Fail-safe mechanisms ensure that vehicles can recover from interrupted or corrupted updates without compromising essential functions. These systems must balance security requirements with availability considerations, ensuring that critical updates can be deployed promptly when addressing discovered vulnerabilities.

5. Advanced Analytics and AI Applications

The integration of advanced analytics and artificial intelligence into vehicular data ecosystems transforms raw telemetry into actionable insights and enables increasingly autonomous vehicle capabilities. Modern vehicles leverage sophisticated AI techniques to improve safety, optimize performance, and enhance the overall driving experience. This section explores the cutting-edge approaches that enable intelligent data utilization across the vehicular ecosystem.

5.1. Federated Learning Approaches for Distributed Model Training

Federated learning has emerged as a transformative paradigm for training machine learning models across distributed automotive systems without centralizing sensitive data. As M. Victoria Luzón, et al. detail in their comprehensive tutorial, this approach enables vehicles to collaboratively improve shared models while keeping raw data securely on local devices [9]. Each vehicle trains models using its local data, sharing only model updates rather than raw training examples. This methodology preserves privacy while enabling fleet-wide learning from diverse driving conditions and scenarios. The federated approach addresses key challenges in vehicular AI development, including data sovereignty concerns, heterogeneous computing environments, and intermittent connectivity that would otherwise impede traditional centralized learning paradigms.

Table 3 Federated Learning Approaches [9, 10]

Approach	Communication Pattern	Aggregation Method	Vehicular Application
Centralized	Vehicle-to-Cloud	Parameter averaging	Fleet-wide optimization
Decentralized	Vehicle-to-Vehicle	Peer-to-peer consensus	Local traffic pattern learning
Hierarchical	Vehicle-to-Edge-to-Cloud	Multi-level aggregation	Regional behavior modeling
Cross-Silo	OEM-to-OEM collaboration	Secure multi-party computation	Industry-wide safety improvements

5.2. AI-Driven Anomaly Detection in Vehicular Systems

The complexity of modern vehicle systems necessitates sophisticated anomaly detection capabilities that can identify potential issues before they manifest as failures. According to Dinh C. Nguyen, et al., AI-driven approaches significantly outperform rule-based methods in detecting subtle deviations from normal operating patterns across interconnected vehicular subsystems [10]. These techniques leverage multimodal sensor data to establish baseline performance models for individual vehicles, accounting for variations in driving conditions, usage patterns, and environmental factors. Unsupervised learning approaches identify emerging anomalies without requiring extensive labeled examples of failure modes, while semi-supervised techniques incorporate domain expertise to prioritize anomalies based on severity and actionability. These systems form a critical component of proactive maintenance strategies and contribute to overall vehicle safety.

5.3. Predictive Maintenance Through Telemetry Analysis

Vehicular telemetry provides rich data streams that enable the transition from scheduled to predictive maintenance paradigms. By analyzing patterns in component behavior over time, AI systems can forecast potential failures before they occur, allowing for proactive intervention. Luzón, et al. highlight how deep learning approaches can identify complex patterns in multivariate time series data that precede component degradation, enabling maintenance scheduling that minimizes downtime while maximizing component lifespans [9]. These predictive systems incorporate contextual factors such as environmental conditions, usage patterns, and maintenance history to refine forecasting accuracy. The resulting maintenance optimization reduces operational costs while improving vehicle reliability and safety through the prevention of unexpected failures.

5.4. Traffic Pattern Optimization Using Aggregated Data

The collective intelligence derived from aggregated vehicular data enables significant improvements in traffic management and routing efficiency. Nguyen, et al. demonstrate how anonymized and aggregated mobility data can inform dynamic traffic optimization systems that reduce congestion, minimize travel times, and lower emissions [10]. These systems leverage reinforcement learning approaches to develop adaptive traffic signal timing strategies that respond to real-time conditions rather than following static schedules. Distributed edge computing nodes process local traffic patterns while contributing to broader network optimization through collaborative learning frameworks. The resulting traffic management systems deliver substantial improvements in urban mobility while reducing environmental impact through more efficient vehicle operation.

5.5. Autonomous Driving Improvements Through Collaborative Learning

The evolution toward fully autonomous vehicles benefits significantly from collaborative learning approaches that accelerate capability development across diverse operating conditions. As detailed by Luzón, et al., federated learning enables autonomous systems to collectively improve perception, prediction, and planning capabilities by sharing insights derived from unique driving scenarios [9]. This collaborative approach allows individual vehicles to benefit from edge cases encountered by the entire fleet without requiring direct experience with rare events. Privacy-preserving techniques ensure that sensitive information about specific journeys remains protected while still contributing to model improvement. The resulting acceleration in autonomous capability development addresses key challenges in achieving safe, reliable autonomous operation across diverse environments and driving conditions.

6. Conclusion

The multifaceted challenges and solutions associated with managing data from cars at scale evolve as vehicles transform into sophisticated mobile data centers. The architectural frameworks span from edge computing paradigms for in-vehicle processing to cloud infrastructure integration, establishing a foundation for efficient data handling across the automotive ecosystem. A critical balance exists between real-time processing requirements and comprehensive analytics capabilities while addressing mobile environment constraints, including bandwidth limitations and storage optimization. Security and privacy frameworks provide essential protection mechanisms ensuring regulatory compliance while safeguarding sensitive vehicular data through robust authentication, threat mitigation, and secure update processes. Advanced analytics and AI applications, particularly federated learning approaches, enable distributed intelligence that preserves privacy while improving vehicle capabilities through collaborative model refinement. As the automotive industry continues transformation toward increasingly connected and autonomous systems, these data management strategies play a pivotal role in realizing the full potential of intelligent transportation networks. Future directions should focus on further optimizing the edge-cloud continuum, enhancing privacy-preserving techniques, and developing more efficient federated learning algorithms specifically tailored to the dynamic and heterogeneous nature of vehicular environments.

References

[1] Wenchao Xu, et al. "Internet of Vehicles in Big Data Era." IEEE/CAA Journal of Automatica Sinica, Vol. 5, No. 1, pp. 19-35, January 2018. DOI: 10.1109/JAS.2017.7510736. https://ieee-jas.net/en/article/doi/10.1109/JAS.2017.7510736

[2] Tom Wong, et al. "PCI Express® Technology: Accelerating Automotive Connectivity, from Infotainment to ADAS." Cadence DriveWorld Conference, August 16-18, 2021. https://pcisig.com/sites/default/files/files/DriveWorld%20PCI-SIG%20-%20Accelerating%20Automotive%20Connectivity%2C%20from%20Infotainment%20to%20ADAS.pdf

[3] Herbert Raj, et al. "A Novel Device-Based Edge-Cloud Architecture for Vehicular Edge Computing." Lecture Notes in Electrical Engineering, Vol. 1274, pp. 1144–1155, October 2024. https://link.springer.com/chapter/10.1007/978-981-97-8043-3_175

[4] Sachin B. Chougule. "Exploring Computing Paradigms for Electric Vehicles: From Cloud to Edge Intelligence, Challenges and Future Directions." World Electric Vehicle Journal, Vol. 15, Issue 2, pp. 39, January 2024. https://www.mdpi.com/2032-6653/15/2/39

[5] Andersen, G., & MoldStud Research Team. "Overcoming Common Challenges in Live Data Processing: Effective Solutions." MoldStud Research Journal, May 3, 2025. https://moldstud.com/articles/p-overcoming-common-challenges-in-live-data-processing-effective-solutions

[6] Yaping Zhang. "Managing Data Bottlenecks: Strategies for Efficient Data Flow across Bandwidth, Storage, and Processing." Hilaris Open Access Journal, 2025 Edition. https://www.hilarispublisher.com/open-access/managing-data-bottlenecks-strategies-for-efficient-data-flow-across-bandwidth-storage-and-processing.pdf

[7] Compliance Hub. "Privacy Laws Compared: CCPA, GDPR, and LGPD Compliance Requirements (2025 Update)." Compliance Hub Wiki, February 15, 2025. https://www.compliancehub.wiki/privacy-laws-compared-ccpa-gdpr-and-lgpd-compliance-requirements-2025-update/

[8] Ulf Mattsson, MSE. "Practical Data Security and Privacy for GDPR and CCPA." ISACA Journal, Volume 3, May 13, 2020. https://www.isaca.org/resources/isaca-journal/issues/2020/volume-3/practical-data-security-and-privacy-for-gdpr-and-ccpa

[9] M. Victoria Luzón, et al. "A Tutorial on Federated Learning from Theory to Practice: Foundations, Software Frameworks, Exemplary Use Cases, and Selected Trends." IEEE/CAA Journal of Automatica Sinica, Vol. 11, No. 4, pp. 824–850, April 2024. DOI: 10.1109/JAS.2024.124215. https://www.ieee-jas.net/article/doi/10.1109/JAS.2024.124215

[10] Dinh C. Nguyen, et al. "Federated Learning for Internet of Things: A Comprehensive Survey." IEEE Transactions on Neural Networks and Learning Systems, 2024 Edition. https://ieeexplore.ieee.org/ielaam/9739/9520094/9415623-aam.pdf