Scalable and Efficient PureChain-Based Federated Learning for Intelligent Transportation Systems

Miraculous Udurume [1], Love Allen Chijioke Ahakonye [2], Jae Min Lee [1], Dong-Seong Kim [1] *
[1] IT-Convergence Engineering, *Kumoh National Institute of Technology*, Gumi, South Korea
* NSLab Co. Ltd., Gumi, South Korea, *Kumoh National Institute of Technology*, Gumi, South Korea
[2] ICT Convergence Research Center, *Kumoh National of Technology*, Gumi, South Korea
(udurumemiraculous@gmail.com, (loveahakonye, ljmpaul, dskim)@kumoh.ac.kr

***Abstract*—Intelligent Transportation Systems (ITS) increasingly leverage edge intelligence for real-time, privacy-preserving, and safety-critical decision-making. However, conventional Federated Learning (FL) frameworks face challenges such as unverifiable updates, central points of failure, and limited scalability. To overcome these issues, we propose PureChain, a scalable FL framework integrated with a permissioned blockchain to enable secure, auditable, and decentralized learning across connected vehicles and road infrastructure. Smart contracts and blockchain validation ensure traceability and integrity of model updates. We simulated the system using three benchmark datasets (BurST-ADMA, Veremi, CICIoV2024), analyzing how variations in client participation and training rounds influence training efficiency and convergence behavior. The results showed a strong linear relationship between the number of clients and sample volume. Extended training improves convergence stability, even in dynamic environments. These findings affirm PureChain's suitability for Software-Defined ITS (SD-ITS), offering both architectural resilience and practical insights for optimizing training configurations in real-world vehicular networks.**

***Index Terms*—Blockchain, Data Efficiency, Edge AI, Federated Learning, Intelligent Transportation Systems, PureChain, Scalability.**

I. Introduction

The advancement of urban mobility requires transportation systems that are intelligent, secure, scalable, and sustainable. Intelligent Transportation Systems (ITS) have emerged as a critical component in achieving this goal by integrating communication, control, and information technologies into vehicles and transportation infrastructure [1]. These systems improve safety, reduce congestion, and improve environmental sustainability, making them essential for modern cities. In recent years, ITS architectures have evolved to incorporate advanced capabilities, including real-time data exchange, edge intelligence, and adaptive control. Technologies such as vehicular-to-everything (V2X) communication, roadside units (RSUs), and vehicular ad hoc networks (VANETs) facilitate distributed intelligence in dynamic traffic environments [2]. With the emergence of edge computing and the anticipated shift towards 6G networks, ITS is increasingly becoming software-defined and data-driven [3]. By giving ITS the ability to make decisions independently and utilize predictive analytics, this change significantly enhances traffic control and accident avoidance.

Federated learning (FL) has emerged as a powerful approach to collaboratively train models at distributed vehicular nodes without compromising data privacy [4]. It enables localized model updates while avoiding the need to transmit raw data, which is especially vital in privacy-sensitive and bandwidth-constrained vehicular environments. Advanced FL mechanisms, including asynchronous updates, intelligent client selection, and edge-aware aggregation [5], further improve performance in mobile and latency-sensitive scenarios. However, ensuring the integrity, traceability, and auditability of federated updates across heterogeneous nodes presents challenges. To address this, blockchain technology has been integrated into FL frameworks, enabling decentralized trust through immutable records and smart contract-based control [6], [7]. This integration is particularly effective in Software-Defined Intelligent Transportation Systems (SD-ITS), where stakeholders operate with varying levels of trust, and the system must withstand adversarial conditions.

Despite recent advances in FL for vehicular networks, a critical gap remains, the impact of participation parameters such as client count and training rounds on data efficiency, consistency, and scalability is underexplored. Although blockchain-integrated FL frameworks address trust and security [7], [8], few offer empirical benchmarks on the participation scale in ITS datasets. This limits the deployment of efficient FL systems in dynamic and resource-constrained environments. To bridge this gap, we introduce PureChain-FL, a blockchain-assisted FL framework that supports decentralized coordination, auditability, and adaptive participation. Our evaluation reveals how participation settings influence convergence and utility in real-world ITS scenarios. The significant contributions of this paper are:

1) We design a system-level architecture for PureChain-FL suited for secure, scalable, and decentralized model training in SD-ITS.
2) We conducted simulation-based evaluations to assess how varying participation parameters affect data collection, model convergence, and system performance.
3) We validate the proposed framework on three benchmark vehicular datasets: BurST-ADMA, Veremi, and CICIoV2024, demonstrating generalizability under various ITS conditions.

The remainder of this paper is structured as follows: Section II presents related works on FL and blockchain integration in ITS. Section III details the architecture and operational flow of the proposed PureChain-FL system. Section IV discusses the experimental setup, the evaluation results, and the performance analysis. Finally, Section V concludes the paper and outlines future research directions.

II. Background and Related Studies

FL has garnered significant attention as a privacy-preserving machine learning paradigm that enables model training in distributed clients without centralizing sensitive data [9]. By transmitting only model updates while retaining data on local edge devices, FL offers strong data privacy guarantees. However, these benefits are often offset by persistent real-world challenges, especially in edge and vehicular environments [5]. Notable issues include communication inefficiencies, statistical heterogeneity, sporadic connectivity, and inconsistent client participation. These factors undermine convergence and model accuracy in highly dynamic contexts such as vehicular networks, where node mobility exacerbates instability. To overcome these challenges, extensive research has introduced solutions such as zone-based aggregation [4], asynchronous and adaptive update strategies [5], and participation-aware optimization techniques that account for client variability and dynamic system conditions. However, the stochastic nature of client availability and sparse data distributions continues to affect the stability of convergence [4]. Furthermore, communication and synchronization bottlenecks hinder the practical deployment of FL in latency-sensitive ITS environments. These challenges highlight the need for an additional trust layer that supports accountability, resilience, and traceability without compromising performance.

Blockchain technology offers a resilient, decentralized trust layer for FL, ensuring tamper-resistant and verifiable model updates through immutability, consensus, and auditability [8]. Moreover, low-latency blockchain frameworks have demonstrated their viability in real-time ITS applications, where responsiveness is crucial [7]. Smart contracts further enhance the ecosystem by enabling secure model aggregation, automated participant authentication, and the enforcement of protocol rules [6]. In IoT-based vehicular ecosystems, distributed consensus and auditability mechanisms have proven effective in managing data integrity across multi-stakeholder infrastructures [10]. Despite these advancements, FL systems remain susceptible to a range of adversarial threats, including poisoning and model inversion attacks [5]. Research has proposed mitigation solutions such as safe multiparty computation, homomorphic encryption, and reputation-based smart contracts to combat these issues [6]. However, a critical research gap remains underexplored: the impact of FL participation dynamics, specifically the number of clients and training rounds, on convergence efficiency, accuracy, and communication cost within mobility-constrained environments [11]. Although participation variability is known to influence model performance, few empirical studies offer benchmarks or frameworks that quantify these effects under realistic vehicular conditions.

To address this gap, PureChain [12] is proposed; A blockchain-enabled FL framework tailored for secure and scalable deployment in SD-ITS. Leveraging PoA^2 consensus mechanism, it integrates decentralized intrusion detection with federated model updates, enabling tamper-proof logging, fault tolerance, and privacy preservation. Unlike general-purpose blockchains, PureChain employs lightweight smart contracts for autonomous anomaly detection, update verification, and trust coordination. Optimized for resource-constrained environments [12], it supports verifiable collaboration without raw data exchange. By evaluating participation heterogeneity, PureChain enhances scalability and efficiency in vehicular FL applications.

III. System Methodology

The proposed PureChain-FL methodology is designed to address the challenges of scalability, security, and data efficiency in SD-ITS. As illustrated in Figure 1, the PureChain architecture features a three-tier design that integrates FL with a permissioned blockchain network to support verifiable collaborative learning that preserves privacy.

A. Blockchain Layer

The blockchain layer ensures governance, trust, and transparency in FL through three core components: smart contracts for decentralized process automation, update validation to enforce the quality and integrity of model updates, and an immutable audit trail for traceability and compliance. Algorithm 1 describes the entire PureChain workflow, where smart contracts automate client coordination, PoA^2 consensus ensures authenticated and context-aware validation of model updates, and secure aggregation enables privacy-preserving, decentralized learning across dynamic SD-ITS environments.

B. PureChain Federated Learning Process

The PureChain-FL process defines a five-stage workflow executed over T communication rounds. At the beginning of each round t, the orchestrator selects a subset of eligible vehicles or edge clients, denoted $C_t \subseteq \{1, ..., N\}$, from the total pool of participants N. Each client $i \in C_t$ downloads the global model M_t and performs local training on its private dataset D^i, yielding an updated model M_t^i in Equation 1.

$$M_t^i = \text{Train}(M_t, D^i). \tag{1}$$

The client then computes its local model update as in Equation 2.

$$\Delta M_t^i = M_t^i - M_t. \tag{2}$$

This update is submitted to the blockchain, forming a transaction $\text{Tx}_i^t = \left(\Delta M_t^i, \text{Sig}_i, \text{Meta}_i\right)$, where Sig_i and Meta_i represent cryptographic and contextual metadata for verification under the PoA^2 consensus mechanism. Only validated updates

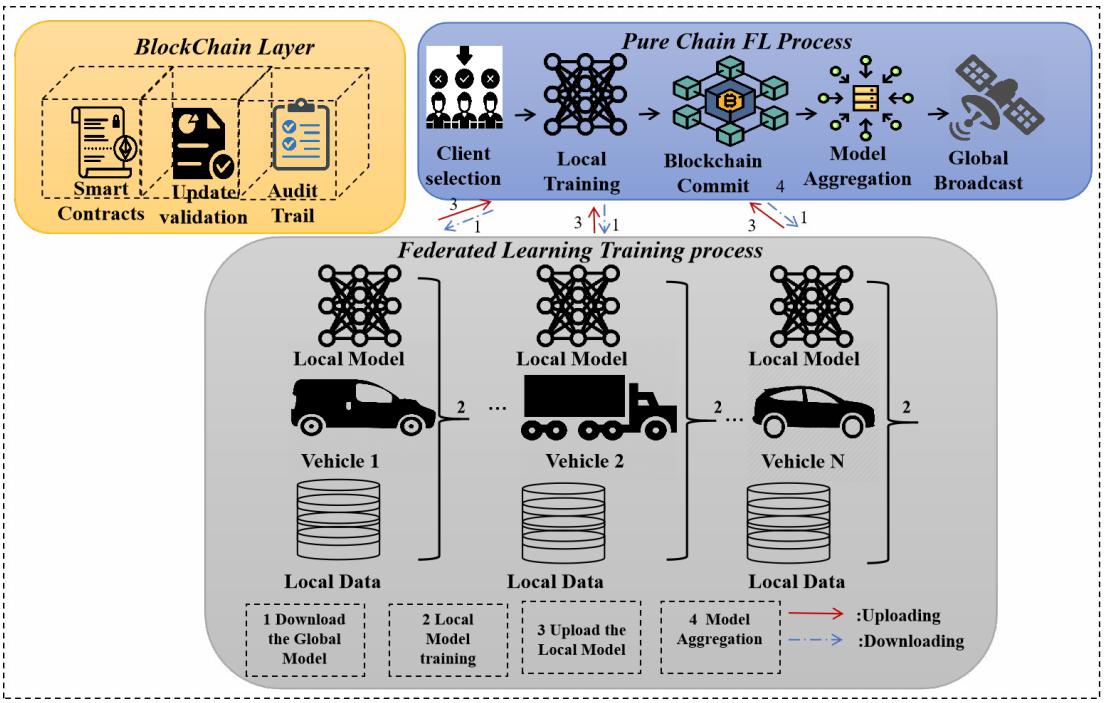


Fig. 1: Overview of PureChain architecture.

Algorithm 1 PureChain Federated Learning with PoA^2 Consensus and Smart Contract Validation

Initialize global model M_0
for each round $t = 1, 2, ..., T$ **do**
  Select subset of clients C_t from total clients N via smart contract
  for each client $i \in C_t$ in parallel **do**
    Train local model M_t^i on private data D^i
    Compute update $\Delta M_t^i = M_t^i - M_{t-1}$
    Sign and submit transaction $\text{Tx}_i^t = (\Delta M_t^i, \text{Sig}_i, \text{Meta}_i)$ to blockchain
  end for
  Validate updates on-chain:
  for each Tx_i^t **do**
    Verify digital signature Sig_i
    Compute association score $\rho_i = \text{Sim}(\phi(D^i), \phi_{\text{global}})$
    Accept ΔM_t^i if $\rho_i \geq \rho_{\min}$
  end for
  Aggregate accepted updates:

$$M_t \leftarrow M_{t-1} + \frac{1}{|\mathcal{A}_t|} \sum_{i \in \mathcal{A}_t} \Delta M_t^i$$

  Broadcast new global model M_t to all clients
end for

$\Delta M_t^i \in \mathcal{A}_t$ are aggregated to form the new global model using federated averaging in Equation 3.

$$M_{t+1} = M_t + \frac{1}{|\mathcal{A}_t|} \sum_{i \in \mathcal{A}_t} \Delta M_t^i. \tag{3}$$

This updated model M_{t+1} is broadcast to all participants, marking the end of the round. The process is repeated for T rounds until the global model converges. Integration of PureChain with PoA^2 validation ensures secure participation, data integrity, and reliable collaboration in vehicular edge environments.

C. Federated Learning Training Process

The third layer of the architecture illustrates the distributed training pipeline in a fleet of vehicles (*Vehicle 1, Vehicle 2, ..., Vehicle N*). Each vehicle maintains a local neural network model, private storage for sensor or operational data, and the ability to download and synchronize with the global model. In each communication round t, the central aggregator initializes the global model M_t by continuing from the results of the previous round in Equation 4.

$$M_t \leftarrow \text{Initialize()} \text{ or } M_{t-1}. \tag{4}$$

Each vehicle $i \in \{1, 2, ..., N\}$ receives M_t and trains a local copy M_t^i using its private data set D^i for epochs E as in Equation 5.

$$M_t^i = \text{Train}(M_t, D^i, E). \tag{5}$$

Upon completion, each client computes the update of the model by subtracting the global model from the local model trained in Equation 6.

$$\Delta M_t^i = M_t^i - M_t. \tag{6}$$

This local update ΔM_t^i is submitted to PureChain for tamper-proof log-in and validation via smart contracts. PureChain ensures the accuracy of the update and the identity of the

contributor. Once all valid updates from the selected client subset C_t are collected, the aggregator performs a weighted averaging to refine the global model in Equation 7.

$$M_{t+1} = M_t + \frac{1}{|C_t|} \sum_{i \in C_t} \Delta M_t^i. \tag{7}$$

This federated training cycle repeats for T communication rounds. With blockchain infrastructure providing cryptographic guarantees, the framework ensures that model updates are verifiable, auditable, and privacy-preserving, enabling secure and decentralized learning within intelligent transportation systems.

Algorithm 2 summarizes the PureChain-FL process, which integrates blockchain with federated model training to ensure secure aggregation, tamper-proof logging, and decentralized coordination between SD-ITS clients.

Algorithm 2 PureChain Federated Learning Process

Initialize global model M_0
for each round $t = 1, 2, ..., T$ **do**
  Select subset of clients C_t from total clients N
  for each client $i \in C_t$ in parallel **do**
    Train local model M_t^i on local data D^i
    Generate update $\Delta M_t^i = M_t^i - M_{t-1}$
    Sign and submit ΔM_t^i to blockchain smart contract
  end for
  Retrieve and verify updates $\Delta M_t^i i = 1^{|C_t|}$ from blockchain
  Aggregate updates: $M_t \leftarrow M_{t-1} + \frac{1}{|C_t|} \sum_i \Delta M_t^i$
  Broadcast M_t to all clients
end for

IV. Experimentation and Result Discussion

A. Dataset Description

This study uses three benchmark data sets: BurST-ADMA [13], Veremi [14] and CICIoV2024 [15], to evaluate the proposed PureChain-FL framework. These datasets were selected for their relevance in simulating real-world vehicle and IoT communication scenarios. The BurST-ADMA dataset features industrial attack simulations and wireless traffic, while the Veremi dataset captures vehicular misbehavior in V2X communications. The CICIoV2024 dataset includes smart city telemetry and threat activities. Each data set varies in size and structure, allowing us to test the scalability and generalization of the framework. All experiments were conducted using Google Colab with Python 3.11.13, TensorFlow Federated, and Matplotlib for visualization. The compute environment was equipped with an Intel(R) Xeon (R) CPU @ 2.20HZ with 12.7 GB RAM running Windows 11. Blockchain operations were simulated using a custom Python-based class, where encrypted client data was appended as blocks during each round of federated learning.

B. Performance Evaluation

The framework was evaluated across different configurations of federated clients and training rounds. For each dataset, experiments were conducted at 5, 10, and 20 clients across 10, 20, and 30 training rounds. Each client contributed fixed local samples (e.g., 100 per round). The primary metric was the cumulative number of training samples collected. Figure 2(a) confirms that PureChain-FL scales efficiently with increasing clients and rounds, ensuring effective data collection for dynamic ITS environments. Figure 2(b) shows the system maintains low-latency inference, with a notable drop at 30 rounds for 10 clients, suggesting model convergence or optimized aggregation. Overall, PureChain-FL supports real-time, decentralized, and scalable learning, making it well-suited for intelligent transportation systems.

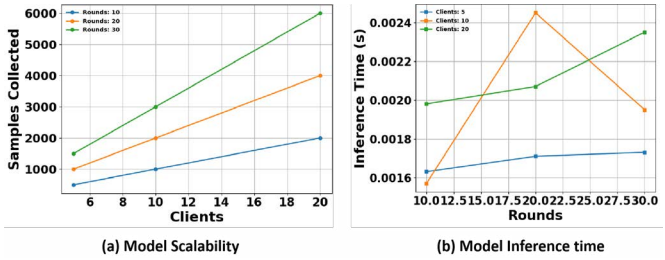


Fig. 2: Impact of sample collection on clients and rounds in BurST-ADMA dataset.

Figure 3(a) demonstrates PureChain-FL's efficient and scalable data collection using the CICIoV2024 dataset, with no signs of performance degradation as clients and rounds increase, supporting large-scale, decentralized ITS environments. Figure 3(b) confirms sustained minimal latency inference between $0.0014s$ and $0.0015s$, with moderate overhead at higher client counts (20), validating the system's suitability for real-time, safety-critical applications.

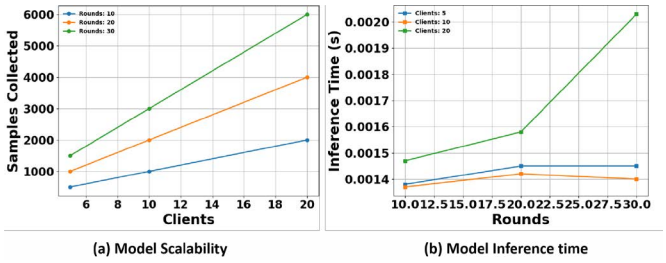
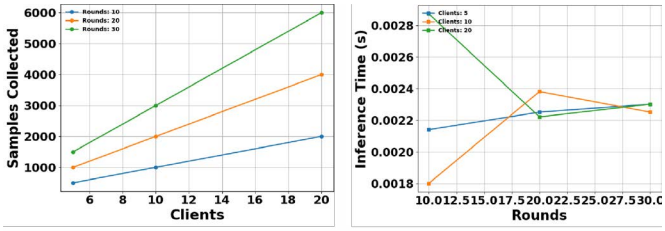


Fig. 3: Impact of sample collection on clients and rounds in CICIoV2024 dataset.

Figure 4(a) confirms PureChain-FL's scalability, with linear sample growth across clients, eliminating aggregation bottlenecks in the Veremi data scenario, which is ideal for dynamic ITS. Figure 4(b) illustrates consistently low inference latency (0.0022–$0.0028s$) across rounds and clients, supporting real-time, safety-critical applications. The system maintains its performance at scale, reflecting the robustness of its blockchain-based, decentralized architecture. Overall, the results highlight

the robustness and scalability of the blockchain-based FL architecture for dynamic and distributed intelligent transportation systems.



(a) Model Scalability (b) Model Inference time

Fig. 4: Impact of sample collection on clients and rounds in Veremi dataset.

The results demonstrate that increasing client participation boosts training sample volume, accelerating model convergence, while more communication rounds enhance model accuracy through abundant aggregation. These findings validate the suitability of PureChain-FL for large-scale and resource-constrained SD-ITS, affirming the effectiveness of blockchain-integrated FL in building robust, trustworthy AI for intelligent transportation systems.

TABLE I: Comparative Analysis of Blockchain-FL Frameworks in ITS

Ref.	Year	FL Type	Blockchain	Security	Scalability
[8]	2020	AV-FL	✓	✓	✗
[7]	2024	FL + IDS	✓	✓	✗
This Study	2025	PureChain-FL	✓	✓	✓

Table I summarizes previous Blockchain-FL frameworks in ITS. Although earlier approaches focus on blockchain-enabled security, they offer limited support for scalability. In contrast, PureChain-FL integrates blockchain technology, enhances security, and enables scalable deployment in dynamic vehicular environments.

V. Conclusions

The experimental results confirm the scalability and data efficiency of the PureChain-based FL for SD-ITS. As client numbers and training rounds increased, training samples scaled linearly across BurST-ADMA, Veremi, and CICIoV2024 data scenarios, demonstrating improved data utilization with broader participation. PureChain's integration of blockchain with FL provides transparency, decentralized trust, and resilience to central failures. The system maintained stable performance and data growth despite fluctuating client availability, crucial for vehicular edge settings. Consistent sample collection across diverse traffic conditions highlights PureChain-FL's robustness to data heterogeneity and its applicability to real-world ITS scenarios. The framework enables secure, verifiable, and privacy-preserving intelligence at the edge. Future work will evaluate metrics such as accuracy, latency, and communication cost, and investigate lightweight consensus, automated smart contracts, and adaptive client sampling for improved efficiency in constrained environments.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government (MSIT) (IITP-2025-RS-2020-II201612, 33%) and by Priority Research Centers Program through the NRF funded by the MEST (2018R1A6A1A03024003, 33%) and by the MSIT, Korea, under the ITRC support program (IITP-2025-RS-2024-00438430, 34%).

References

[1] H. L. Nakayiza, L. A. C. Ahakonye, D.-S. Kim, and J. M. Lee, "Blockchain-Enhanced Feature Engineered Data Falsification Detection in 6G In-Vehicle Networks," *IEEE Internet of Things Journal*, pp. 1–1, 2025.

[2] W. Albattah, S. Habib, M. F. Alsharekh, M. Islam, S. Albahli, and D. A. Dewi, "An Overview of the Current Challenges, Trends, and Protocols in the Field of Vehicular Communication," *Electronics*, vol. 11, no. 21, p. 3581, 2022.

[3] T. Alqubaysi, A. F. A. Asmari, F. Alanazi, A. Almutairi, and A. Armaghan, "Federated Learning-Based Predictive Traffic Management Using a Contained Privacy-Preserving Scheme for Autonomous Vehicles," *Sensors*, vol. 25, no. 4, p. 1116, 2025.

[4] J. Sam, A. Kumar, and P. Singh, "A Survey of Federated Learning in Vehicular Networks: Challenges, Architectures, and Applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 5, pp. 4150–4165, 2022.

[5] M. Rezaei, A. Sabzalian, H. Karimipour, Z. Lin, and I. Khalil, "Fed-VT: Federated vehicular-tier learning for autonomous driving systems," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[6] S. N. Khan, F. Loukil, C. Ghedira-Guegan, E. Benkhelifa, and A. Bani-Hani, "Blockchain smart contracts: Applications, challenges, and future trends," *Peer-to-peer Networking and Applications*, vol. 14, no. 5, pp. 2901–2925, 2021.

[7] Z. Abou El Houda, H. Moudoud, B. Brik, and L. Khoukhi, "Blockchain-enabled federated learning for enhanced collaborative intrusion detection in vehicular edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 7, pp. 7661–7672, 2024.

[8] S. R. Pokhrel and J. Choi, "Federated Learning with Blockchain for Autonomous Vehicles: Analysis and Design Challenges," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4734–4746, 2020.

[9] D. Shenoy, R. Bhat, and K. Krishna Prakasha, "Exploring privacy mechanisms and metrics in federated learning," *Artificial Intelligence Review*, vol. 58, p. 223, 2025.

[10] L. Ahakonye, C. Nwakanma, and D.-S. Kim, "Tides of Blockchain in IoT Cybersecurity," *Sensors*, vol. 24, p. 3111, 05 2024.

[11] M. Udurume, J. N. Njoku, V. Shakhov, and I. Koo, "A Review of Intrusion Detection Techniques in MQTT-Enabled IoT Network," *IEEE Access*, pp. 1240–1245, 2024.

[12] D.-S. Kim, I. S. Igboanusi, L. A. Chijioke Ahakonye, and G. O. Anyanwu, "Proof-of-Authority-and-Association Consensus Algorithm for IoT Blockchain Networks," in *2025 IEEE International Conference on Consumer Electronics (ICCE)*, 2025, pp. 1–6.

[13] M. A. Amanullah, M. Baruwal Chhetri, S. W. Loke, and R. Doss, "BurST-ADMA: Towards an Australian Dataset for Misbehaviour Detection in the Internet of Vehicles," in *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, 2022, pp. 624–629.

[14] J. Kamel, M. Wolf, R. W. van der Hei, A. Kaiser, P. Urien, and F. Kargl, "VeReMi Extension: A Dataset for Comparable Evaluation of Misbehavior Detection in VANETs," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[15] E. C. P. Neto, H. Taslimasa, S. Dadkhah, S. Iqbal, P. Xiong, T. Rahman, and A. A. Ghorbani, "CICIoV2024: Advancing Realistic IDS Approaches Against DoS and Spoofing Attack in IoV CAN Bus," *Internet of Things*, vol. 26, p. 101209, 2024.