A Meta Learning Framework for Intrusion Detection in Connected Vehicles

Okba BEN ATIA[a], Mustafa AL SAMARA[b], Ismail BENNIS[b]
[a]Université Marie et Louis Pasteur, UTBM, CNRS, Institut FEMTO-ST
F-90000 Belfort, France
[b]IRIMAS, University of Haute-Alsace, France
{mustafa.al-samara, ismail.bennis}@uha.fr, okba.ben-atia@utbm.fr

***Abstract*—The Controller Area Network (CAN) protocol is the main communication backbone in modern vehicles, enabling data exchange between Electronic Control Units (ECUs). However, its lack of built-in security mechanisms exposes Connected Vehicles (CV) to a growing range of cyber threats. To address this, we propose a distributed intrusion detection framework that integrates Model-Agnostic Meta-Learning (MAML) with Federated Learning (FL), offering adaptability and privacy preservation. The system leverages an LSTM-based model to learn temporal patterns in CAN message IDs, enabling the detection of known and previously unseen attacks. MAML enhances the model's generalization by facilitating rapid adaptation to new threat types using limited data, while FL enables collaborative training across multiple CVs without sharing raw data. Our distributed approach ensures continuous learning and robustness in real-world automotive environments. Experimental results demonstrate the framework's effectiveness in detecting complex intrusions while maintaining data privacy across participating vehicles.**

***Index Terms*—Connected Vehicles (CV), Model-Agnostic Meta-Learning (MAML), Federated Learning (FL), Controller Area Network (CAN), Intrusion Detection System (IDS), Long Short-Term Memory (LSTM), Denial-of-Service (DoS) Attack, Fuzzy Attack.**

I. Introduction

The increasing reliance on interconnected systems in Connected Vehicles (CV), including autonomous and non-autonomous vehicles, introduces critical challenges to their security and reliability. The Controller Area Network (CAN) protocol, a lightweight and efficient communication standard, is widely used in modern vehicles to enable seamless interactions between Electronic Control Units (ECU) [1]. However, the CAN protocol lacks robust security mechanisms, making it vulnerable to cyberattacks such as Denial-of-Service (DoS) and Fuzzy attacks [2]. These vulnerabilities pose significant risks to CV safety and reliability, necessitating the development of advanced Intrusion Detection Systems (IDS) tailored to the dynamic and evolving vehicular environment [3]. Traditional IDS approaches often rely on centralized training architectures or static models that fail to adapt to unseen attack patterns [4], [5]. Furthermore, centralized methods raise privacy concerns due to the sensitive nature of vehicular data [6], [7]. To address these limitations, we propose Meta-FLCV, a novel IDS architecture that integrates Model-Agnostic Meta-Learning (MAML) and Federated Learning (FL). Recent surveys have highlighted the potential of FL to improve CV performance and security [8]. The integration of MAML enables the IDS to adapt to novel attack patterns with minimal additional data rapidly. At the same time, FL facilitates collaborative model training across distributed CV without sharing raw data, ensuring privacy.

In this paper, we utilize the Car-Hacking dataset [9], which is based on CAN messages and represents the backbone of communication in CV. The dataset includes benign and malicious traffic, focusing on DoS and Fuzzy attacks that exploit the inherent vulnerabilities of the CAN protocol. Our architecture employs a Long Short-Term Memory (LSTM)-based model to process CAN message IDs, effectively capturing temporal dependencies to classify benign and malicious messages. By leveraging MAML, the IDS achieves exceptional adaptability to unseen attack patterns, while FL enables privacy-preserving collaborative training across multiple CVs, utilizing Federated Averaging for model aggregation.

The main contributions of this paper can be summarized as follows:

- We propose Meta-FLCV, a novel IDS architecture that integrates MAML and FL for intrusion detection in CV, addressing adaptability and privacy challenges.
- We employ an LSTM-based model to process CAN message IDs, leveraging their sequential characteristics to enhance intrusion detection accuracy.
- We utilize MAML to improve the IDS's ability to quickly adapt to novel and evolving attack patterns with minimal additional data.
- We integrate FL to enable collaborative training across distributed CV, preserving privacy by sharing only model parameters instead of raw data.
- We demonstrate the exceptional performance of Meta-FLCV in detecting and adapting to DoS and Fuzzy attacks using the Car-Hacking dataset [9], highlighting its potential as a robust and adaptive

979-8-3503-9281-4/25/$31.00 ©2025 IEEE

IDS for secure CV environments.

The remainder of this paper is organized as follows: Section II reviews related works on traditional vehicular IDS. Section III presents the Meta-FLCV architecture and its components. The evaluation setup and results are detailed in Section IV. Finally, Section V concludes the paper with future research directions.

II. Related Works

IDS are essential for ensuring CV security by mitigating attacks targeting communication protocols like the CAN. Traditional IDS approaches often rely on centralized architectures and static machine learning models, which struggle to adapt to vehicular environments' dynamic and heterogeneous nature. In [4], the authors introduced a Generative Adversarial Network (GAN)-based IDS, known as GIDS, for anomaly detection in in-vehicle networks. While this method demonstrated high accuracy in identifying deviations from normal traffic, its heavy dependence on large labeled datasets limits its flexibility in detecting new attack patterns. Similarly, the authors of [9] proposed a Convolutional Neural Network (CNN) based IDS for CAN message analysis. Although effective in achieving high detection rates, the model's static nature restricts its ability to generalize across varied vehicular environments and evolving threats.

FL has emerged as a promising privacy-preserving approach for distributed data analysis. In [5], the authors explored the use of FL for intrusion detection in IoT networks, highlighting its capability for collaborative training without centralizing data. However, their study did not address the specific challenges of vehicular networks, such as high-frequency sequential data and the low-latency demands of CAN communications.

In adaptive learning, authors of [6] introduced MAML, a technique designed for rapid model adaptation using minimal additional data. Despite its success in various domains, MAML's application in vehicular IDS remains limited. Integrating MAML into IDS frameworks could potentially overcome the limitations of static models like GIDS [4] and CNN-based IDS [9], enabling faster adaptation to emerging attack patterns.

In [7], the authors proposed a hierarchical FL approach for CV, focusing on privacy and efficient collaboration. While their method improved communication efficiency in vehicular networks, it lacked mechanisms for rapid adaptation to evolving cyber threats. Furthermore, authors of [10] developed an IDS leveraging a Deep Convolutional Neural Network (DCNN) for in-vehicle networks, utilizing CAN traffic's temporal and sequential properties. This approach achieved high detection accuracy and low false-negative rates, outperforming traditional machine learning models. However, relying on supervised learning limits its effectiveness in identifying unseen attack types.

In [11], the authors introduced a hybrid method combining FL and adaptive learning for IoT-based IDS. Although their approach highlighted the benefits of privacy-preserving collaborative learning, it lacked optimizations specific to CAN-based vehicular networks, where ID-based sequential data plays a critical role in effective intrusion detection.

Several studies have significantly advanced the field of CV security. However, each has notable limitations. Despite its high anomaly detection accuracy, the GIDS in [4] requires large labeled datasets, limiting its adaptability to novel attacks. The CNN-based IDS in [9] is effective in controlled environments but lacks generalization across diverse vehicular settings. The FL approach in [5] enhances privacy and collaboration but overlooks the unique demands of vehicular networks, such as high-frequency sequential data. Despite its effectiveness for rapid adaptation, MAML [6] is underexplored for vehicular IDS. The hierarchical FL in [7] improves scalability but lacks mechanisms for swift threat adaptation. The hybrid FL and adaptive learning approach of [11] is effective in IoT contexts but not optimized for CAN-based vehicular networks. Moreover, these studies mainly focus on known attacks, with limited attention to rapid adaptation to novel threats using minimal data. These gaps underscore the need for a more adaptive, robust IDS capable of addressing CV security's dynamic and privacy-sensitive landscape.

Building on these studies, we propose Meta-FLCV, an innovative IDS architecture that integrates MAML and FL to address the dual challenges of adaptability and privacy in CV. By leveraging MAML, Meta-FLCV enables rapid model fine-tuning for novel attack scenarios, overcoming the static nature of traditional IDS methods. The integration of FL allows for collaborative training across multiple clients (vehicles) while preserving data privacy, as only model parameters are shared during training. This approach ensures rapid adaptability and robustness, making it a comprehensive solution for intrusion detection in dynamic vehicular environments.

III. Meta-FLCV proposed model

Meta-FLCV is a novel framework designed to improve intrusion detection in vehicular CAN bus by integrating FL and MAML. This model addresses the challenges of detecting malicious intrusions, such as DoS attacks, while ensuring data privacy and enabling rapid adaptation to unseen attacks. The Car-Hacking dataset [9] is utilized to evaluate its effectiveness, containing multiple types of intrusions, including DoS attacks, Fuzzy attacks, and Spoofing attacks. In our approach, we evaluate our model using two distinct scenarios. In the first scenario, we train our model on labeled DoS attack data and test it on the same attack type to measure detection performance in a known attack environment. In the second scenario, we test the model on Fuzzy

attack data, representing an unseen attack scenario, to demonstrate its adaptability to new attack patterns.

As shown in Fig. 1, our model operates through a series of structured steps. First, message monitoring is performed within each vehicle to capture CAN traffic data, including the CAN ID (e.g., '0x123'), payload data ('DATA[0-7]'), and the timestamp of message transmission. These messages are labeled as either normal or malicious based on their characteristics. For instance, normal traffic might include messages like 'CAN ID: 0x123, DATA: 00 01 02 03 04 05 06 07, Timestamp: 1479121434.860229', whereas DoS attack messages might appear as 'CAN ID: 0x000, DATA: 00 00 00 00 00 00 00 00, Timestamp: 1479121434.860700', where high-frequency, repetitive injections with the same CAN ID indicate malicious behavior. To evaluate the effectiveness and adaptability of our framework, we focus on two scenarios: the DoS attack is used as the known threat for training the model, while the Fuzzy attack, characterized by random CAN IDs and payloads, is used as an unseen scenario to test the system's ability to adapt rapidly with minimal additional data.

Fig. 1 also provides a high-level architectural overview of the proposed Meta-FLCV framework. It incorporates common wireless and physical attack vectors to contextualize the threat landscape. The vehicle processing unit is visually emphasized to reflect its central role in real-time detection and model adaptation. The presence of attacker elements is illustrative, intended to indicate potential points of entry rather than actual breaches. In general, the figure emphasizes the solution architecture while situating it within the realistic cybersecurity challenges faced by CVs.

Next, local intrusion detection is performed within each vehicle using an LSTM model with 32 hidden units, which analyzes the temporal patterns in the message sequences. LSTM excels at learning sequential dependencies, making it suitable for identifying anomalies such as high-frequency injections in DoS attacks. The model outputs probabilities indicating whether each message is normal ("0") or malicious ("1"). After detecting anomalies locally, MAML is applied to improve the adaptability of the model. Within each vehicle, the Inner Loop updates the model using the local data to optimize its performance on specific tasks. The Outer Loop, also executed within the vehicle, generalizes the model to ensure robustness across multiple tasks within the same vehicle. FL is then utilized to aggregate the knowledge gained from all vehicles. Each vehicle sends its locally updated model parameters to a central server without sharing raw data, preserving data privacy. The server performs Federated Averaging (FedAvg) to combine the models into a single global model, which is redistributed back to all vehicles. This global model benefits from the collective learning of all vehicles, making it more effective at detecting a wide range of intrusions. Finally, the Meta-FLCV model ensures rapid adaptation to unseen attacks. MAML equips the global model with the capability to generalize across diverse attack types, enabling it to adapt quickly to new attack patterns, such as Fuzzy attacks, with minimal additional training. The Meta-FLCV model's key contributions lie in its ability to combine the strengths of FL and Meta-Learning to achieve a highly adaptable, privacy-preserving, and efficient IDS for distributed vehicular networks. By leveraging temporal pattern recognition and collaborative learning, Meta-FLCV represents a robust solution for securing modern automotive systems against evolving cyber threats.

This process is described in algorithm 1.

Algorithm 1: Meta-FLCV

Input: Car-Hacking Dataset, labeled DoS attack data for training, unseen Fuzzy attack data for testing
Output: Trained and generalized intrusion detection model
Step 1: Data Collection & Preprocessing
Capture and preprocess CAN traffic data
Label messages as normal or malicious based on characteristics
Step 2: Local LSTM-based Intrusion Detection
Train an LSTM model on vehicle-specific data to capture temporal patterns
Step 3: Meta-Learning with MAML
Inner Loop: For each vehicle i, update local parameters: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_i$
Outer Loop: Aggregate adapted parameters of the vehicles: $\theta \leftarrow \theta - \beta \nabla_\theta \sum_i \mathcal{L}_i$
// θ: global model, θ_i': locally adapted model, α: inner learning rate (per vehicle), β: outer learning rate (global update), $\mathcal{L}_i$: local loss
Step 4: Federated Learning Aggregation
Each vehicle sends updated model parameters θ_i' to the central server
Perform Federated Averaging (FedAvg)
Step 5: Evaluation & Comparison
Evaluate model using unseen Fuzzy attack data
Adapt model to new attack patterns with minimal additional training
return *Final optimized Meta-FLCV intrusion detection model*

IV. Experimental results and analysis

A. Dataset

Our experimental analysis is based on the Car-Hacking dataset [9], which includes three types of CAN bus attacks: DoS, Fuzzy, and Spoofing (RPM/Gear). The dataset was collected from a real vehicle by logging CAN traffic via the OBD-II port while message injection attacks were performed. The dataset is publicly available at [12].

For our evaluation, we use the DoS attack as labeled data for training and testing, while the Fuzzy attack is treated as an unseen attack to test our framework under real-world conditions. This setup allows us to evaluate the generalization ability of our model when encountering novel attacks. Table 1 summarizes the characteristics of the dataset used in our experiments.

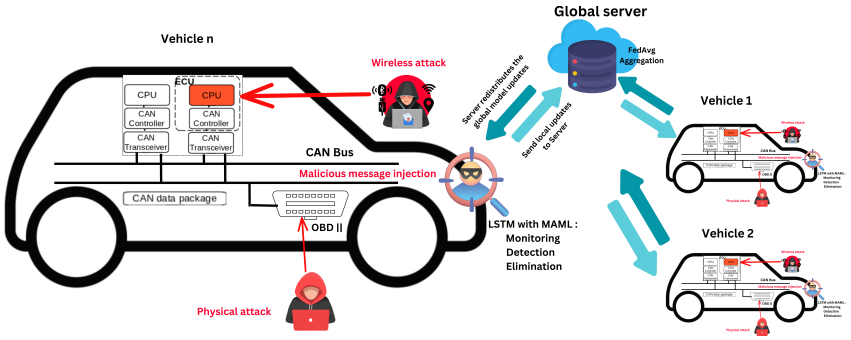

Fig. 1. Meta-FLCV framework

Attack Type	Description	Injection Frequency	CAN ID	DLC Range	Data Field	Flag
DoS Attack (Labeled)	Injecting messages of '0000' CAN ID every 0.3 ms	0.3 ms	0000 (Dominant)	0-8	DATA[0]-DATA[7]	T (Injected)
Fuzzy Attack (Unseen)	Injecting messages of random CAN ID and DATA values every 0.5 ms	0.5 ms	Random	0-8	DATA[0]-DATA[7]	T (Injected)
Spoofing Attack (RPM/Gear)	Injecting messages of specific CAN ID related to RPM/gear information	1.0 ms	Certain CAN IDs	0-8	DATA[0]-DATA[7]	T (Injected)

TABLE 1
CAR-HACKING DATASET CHARACTERISTICS

B. Evaluation metrics

Several metrics for evaluating the performance of Meta-FLCV are used in CV. The first metric commonly used is the accuracy, which measures the overall correctness of the model's predictions. It is defined in Equation (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. However, when dealing with highly imbalanced datasets, as is often the case in CV-based IDS, accuracy alone may not provide a comprehensive evaluation of the model's effectiveness.

We also consider the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) to address this limitation, which is a robust metric for evaluating classification performance, especially in imbalanced datasets. AUC measures the model's ability to distinguish between positive and negative classes and is calculated as in Equation (2):

$$AUC = \int_0^1 TPR(FPR)\, d(FPR) \tag{2}$$

Where the True Positive Rate (TPR) and False Positive Rate (FPR) are defined as in Equation (3):

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN} \tag{3}$$

AUC is critical in Meta-FLCV because it provides an aggregated performance measure across all classification thresholds, making it more reliable for imbalanced datasets. Another essential metric is the Loss function, which quantifies the error between predicted and actual values. In the case of classification, Cross-Entropy Loss is commonly used and defined as in Equation (4):

$$\mathcal{L} = -\sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{4}$$

Where y_i is the true label, $\hat{y}_i$ is the predicted probability, and N is the total number of samples. Loss is crucial for optimizing Meta-FLCV, as it guides the model's learning process by minimizing classification errors. By incorporating these metrics, Meta-FLCV ensures a comprehensive evaluation of the model's effectiveness in CV tasks, addressing overall performance and robustness against imbalanced datasets.

C. Experimental setup

Table 2 presents the key system specifications, model configurations, and training parameters used in the our framework. In this section, we provide the performance analysis of Meta-FLCV with two different scenarios. In the first scenario, Meta-FLCV is evaluated on the DoS attack as a known threat. The AUC results in Fig 2 show consistently high performance (AUC = 100%)

TABLE 2
META-FLCV EXPERIMENTAL SETUP

Component	Details
System	Debian 12, Intel i7-13700H (20 threads, 5.0 GHz), 32GB RAM
GPU	NVIDIA RTX A500 Laptop (4GB VRAM, CUDA 12.4)
Frameworks	PyTorch 2.5, Flower 1.12.0
Model	LSTM (Input: 9 features: 1 CAN ID + 8 data bytes, Hidden size: 32, Output: 1)
Optimization	Outer-loop: Adam (LR 0.001), Inner-loop: SGD (LR 0.05)
Training	20 local epochs, batch size 32
Federation	5 clients, 20 rounds, FedAvg aggregation
Libraries	Pandas 2.1.1, NumPy 1.24.3

for Vehicles 1, 3, and 5, which received shared updates from the global server, while Vehicles 2 and 4 exhibit noticeable fluctuations, with AUC dropping as low as 79% in early epochs before stabilizing. This highlights the role of collaborative updates in ensuring detection stability. Similarly, the loss in Fig 3 remains consistently

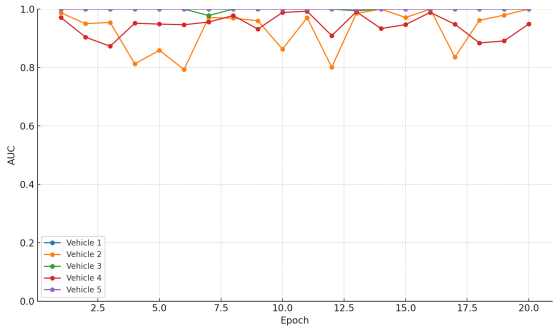

Fig. 2. AUC under DoS attack.

low for Vehicles 1, 3, and 5, averaging around 0.0034, demonstrating efficient training and convergence due to shared updates. In contrast, Vehicles 2 and 4 experience significantly higher loss fluctuations, with values reaching 0.65 at some points. The instability in the loss for these vehicles suggests difficulties in optimization and convergence due to their reliance solely on local updates, further emphasizing the impact of excluding them from the global aggregation process. The lack of knowledge sharing leads to slower learning and less stability in their decision boundaries, making their performance less reliable.

Accuracy in Fig. 4 follows a similar trend, with Vehicles 1, 3, and 5 achieving near-perfect classification performance at approximately 98% throughout training. Their stable accuracy confirms the effectiveness of Meta-FLCV in detecting known attacks when global updates are incorporated. However, Vehicles 2 and 4 show reduced and fluctuating accuracy, as their models struggle to generalize effectively without collaborative learning.

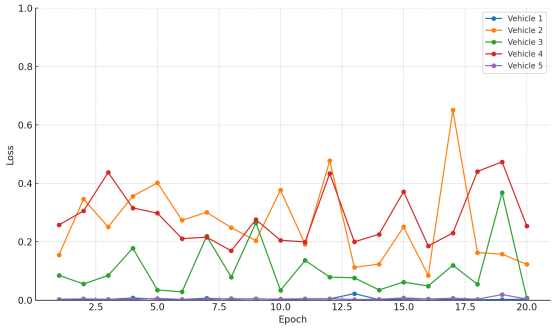

Fig. 3. Loss under DoS attack.

This reduced stability in classification highlights the necessity of shared updates in FL settings, ensuring that all clients benefit from global knowledge to improve robustness and consistency.

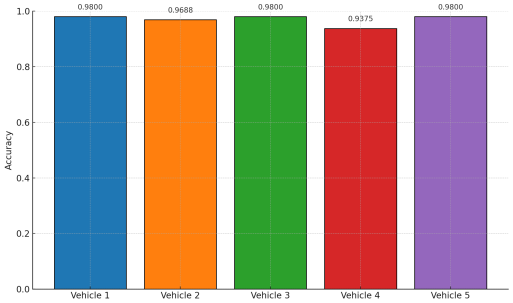

Fig. 4. Accuracy under DoS attack.

In the second scenario, the adaptability of Meta-FLCV is tested on the Fuzzy attack, an unseen threat. The AUC in Fig. 5 demonstrates a significant improvement, starting at 75% and progressively increasing to 99%, reflecting the model's ability to adapt to new attack patterns quickly. The ability to enhance detection performance with minimal exposure to new attack data confirms the effectiveness of Meta-FLCV in real-world dynamic environments.

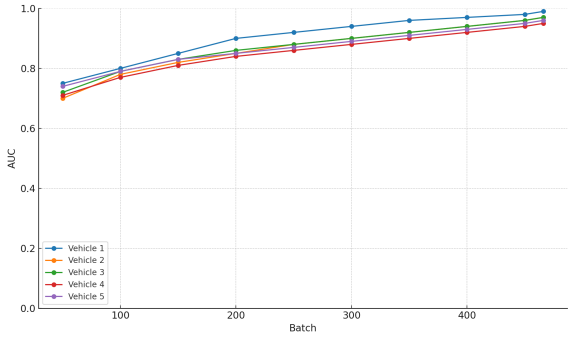

Fig. 5. AUC under Fuzzy attack.

The loss in Fig. 6 follows a steadily decreasing

trend, dropping from 0.85 to 0.08, reinforcing the efficiency of the meta-learning process. This loss reduction suggests that the model successfully incorporates the new knowledge and fine-tunes its decision boundaries with minimal labeled data, ensuring faster adaptation to emerging attacks. The ability to reduce loss efficiently demonstrates the capability of Meta-FLCV to learn from small increments of new attack data without requiring extensive retraining.

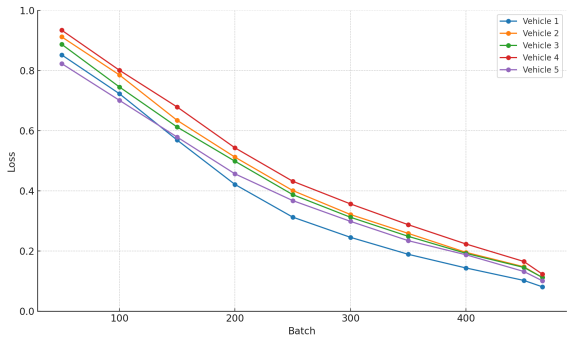

Fig. 6. Loss under Fuzzy attack.

Similarly, accuracy in Fig. 7 improves drastically, rising from 68.75% in the initial phase to 100% at the end of the training process. This significant increase validates the strength of the MAML process, which allows the model to generalize to unseen attacks quickly. The use of small, progressively increasing batches further enhances this adaptability, proving that Meta-FLCV is well-equipped to handle evolving threats in security-critical environments with minimal data availability. We conducted a feature importance analysis that shows that temporal patterns in CAN IDs and specific payload characteristics are key to detecting attacks and enhancing interpretability without compromising accuracy.

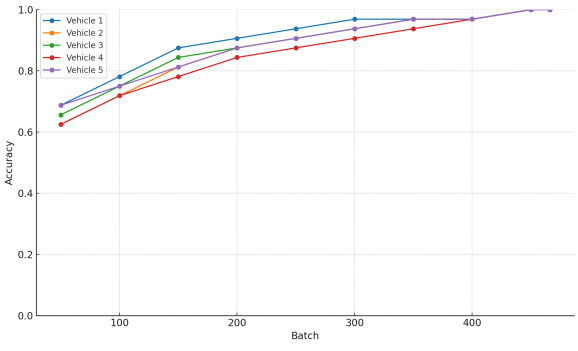

Fig. 7. Accuracy under Fuzzy attack.

In summary, the results confirm that Meta-FLCV achieves strong detection performance in the known attack scenario and exhibits rapid adaptability to unseen attacks. The impact of shared updates is evident, as vehicles without global updates show instability. At the same time, the framework's meta-learning capabilities enable quick generalization to new threats, reinforcing its suitability for real-world deployment in evolving security environments.

V. Conclusion

In this paper, we introduce Meta-FLCV, a privacy-preserving framework for intrusion detection in CAN-based CV. By integrating MAML with FL, the framework enables adaptability to novel attack scenarios while ensuring privacy. Our evaluation showed high detection accuracy in known DoS attacks and strong generalization to unseen Fuzzy attacks demonstrated through a custom scenario simulating varied conditions, highlighting its effectiveness in dynamic vehicular environments.

For future work, we aim to extend Meta-FLCV to broader attack types, ensuring robust and scalable intrusion detection, compare its performance with other approaches, and analyze the impact of communication overhead when scaling to a larger dataset with a bigger number of vehicles, model update latency, and bandwidth requirements, which are crucial for real-time IDS.

References

[1] Jo, H.J. and Choi, W., 2021. A survey of attacks on controller area networks and corresponding countermeasures. IEEE Transactions on Intelligent Transportation Systems, 23(7), pp.6123-6141.

[2] Kim, K., Kim, J.S., Jeong, S., Park, J.H. and Kim, H.K., 2021. Cybersecurity for autonomous vehicles: Review of attacks and defense. Computers & security, 103, p.102150.

[3] Smith, J. A., Brown, L. M., & Davis, K. R., 2023. An Investigation of Cyber-Attacks and Security Mechanisms for Connected and Autonomous Vehicles. *IEEE Transactions on Vehicular Technology*, 72(4), 567-580.

[4] Seo, E., Song, H.M., and Kim, H.K., 2018. GIDS: GAN-based intrusion detection system for in-vehicle network. In 2018 16th Annual Conference on Privacy, Security and Trust (PST), pp. 1–6.

[5] Nguyen, T., Nguyen, G., Phung, D., and Nahavandi, S., 2021. Federated learning for intrusion detection in IoT networks. IEEE Internet of Things Journal, 8(1), pp. 494–504.

[6] Finn, C., Abbeel, P., and Levine, S., 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the 34th International Conference on Machine Learning (ICML), pp. 1126–1135.

[7] Wang, J., Shi, M., Wang, X., and Shen, X.S., 2021. Hierarchical federated learning for connected vehicles. IEEE Transactions on Vehicular Technology, 70(4), pp. 3305–3316.

[8] Chellapandi, V. P., Yuan, L., Brinton, C. G., Zak, S. H., & Wang, Z., 2023. Federated Learning for Connected and Automated Vehicles: A Survey of Existing Approaches and Challenges. *IEEE Transactions on Intelligent Vehicles*, 8(2), 123-145.

[9] Song, H.M., Woo, J., and Kim, H.K., 2020. In-vehicle network intrusion detection using deep convolutional neural network. Vehicular Communications, 21, p.100198.

[10] Song, H.M., Woo, J., and Kim, H.K., 2020. In-vehicle network intrusion detection using deep convolutional neural network. *Vehicular Communications*, 21, p.100198. DOI: https://doi.org/10.1016/j.vehcom.2019.100198.

[11] Zhang, X., Li, Y., and Chen, W., 2023. Hybrid federated and adaptive learning techniques for intrusion detection in IoT systems. *Journal of Internet Security Research*, 12(3), pp. 45–60.

[12] OCS Lab, "Car Hacking Dataset." [Online]. Available: https://ocslab.hksecurity.net/Datasets/car-hacking-dataset [Accessed: June 2025].