Cyber-Social Security for Smart Mobility: Assessing Adversarial Machine Learning Threats in Connected Vehicles

1st Vita Santa Barletta
Department of Computer Science
University of Bari Aldo Moro
Bari, Italy
vita.barletta@uniba.it

2nd Paolo Buono
Department of Computer Science
University of Bari Aldo Moro
Bari, Italy
paolo.buono@uniba.it

3rd Danilo Caivano
Department of Computer Science
University of Bari Aldo Moro
Bari, Italy
danilo.caivano@uniba.it

4rd Antonio Curci
Department of Computer Science
University of Bari Aldo Moro
Bari, Italy
antonio.curci@uniba.it

5th Samuele del Vescovo
Department of Computer Science
University of Bari Aldo Moro
Bari, Italy
samuele.delvescovo@imtlucca.it

6th Antonio Piccinno
Department of Computer Science
University of Bari Aldo Moro
Bari, Italy
antonio.piccinno@uniba.it

***Abstract*—Connected and Autonomous Vehicles (CAVs) are becoming key components of Smart Cities, enabling intelligent transportation systems, enhanced urban mobility, and improved road safety. However, the adoption of AI technologies for vehicle cybersecurity introduces new attack surfaces that may affect the resilience of smart mobility infrastructures. This paper investigates the impact of Black-Box Adversarial Machine Learning (AML) attacks against ML-based Intrusion Detection Systems (IDSs) operating on automotive Controller Area Network (CAN) communications. The study focuses on two representative AML techniques: the Zeroth-Order Optimization (ZOO) evasion attack and the Membership Inference (MI) attack. Experiments were conducted on OTIDS benchmark dataset using several supervised learning algorithms, including Decision Trees, Random Forests, Gradient Boosting, Extreme Gradient Boosting, and K-Nearest Neighbors. Results show that ZOO attacks can significantly degrade Intrusion Detection Systems (IDSs) performance, causing an average reduction of approximately 70% in weighted accuracy, whereas MI attacks proved ineffective under the considered conditions. Beyond the technical evaluation, the paper discusses the implications of AML threats for Smart Mobility and Urban Security, highlighting their potential impact on transportation services, situational awareness, and urban resilience. Furthermore, adversarial training is evaluated as a mitigation strategy and demonstrates its effectiveness in improving IDS robustness. The findings contribute to the design of secure AI-enabled mobility systems and resilient cybersecurity architectures for future Smart Cities.**

***Index Terms*—Autonomous Vehicles, Cyber-Social Security, Smart Mobility, Attacks**

I. Introduction

The adoption of Vehicle-to-Everything (V2X) communications, cloud services, and edge computing is transforming Smart City transportation systems by improving mobility and situational awareness [1], [2]. At the same time, the growing connectivity of vehicles expands the cyber-attack surface of urban infrastructures [3].

To enhance vehicle cybersecurity, Machine-Learning-(ML)-based Intrusion Detection Systems (IDSs) are increasingly deployed to monitor in-vehicle Controller Area Network (CAN) communications. However, these solutions are vulnerable to Adversarial Machine Learning (AML) attacks, which manipulate input data to evade detection or compromise model performance.

In this context, Cyber-Social Security (CSS) provides a multidisciplinary framework that integrates technical, informational, and human-centric dimensions to analyze cyber threats affecting interconnected Smart City infrastructures [4], [5]. As illustrated in Figure 1, the proposed framework links the Smart City ecosystem, the CSS dimensions, and AML threats targeting automotive IDSs, highlighting their potential impact on mobility services, public safety, and urban resilience.

Therefore, this paper investigates the impact of two representative Black-Box AML attacks, namely Zeroth-Order Optimization (ZOO) and Membership Inference (MI), against ML-based IDSs operating on automotive CAN networks. The objective is to evaluate the resilience of AI-enabled vehicle cybersecurity mechanisms and discuss the implications of adversarial attacks for Smart Mobility and Urban Security.

II. Methodology

The experimental framework models a Smart City mobility ecosystem in which connected vehicles rely on Machine Learning-based Intrusion Detection Systems to identify cyber threats affecting in-vehicle communications. The goal is to evaluate the resilience of AI-enabled vehicle cybersecurity mechanisms against Black-Box Adversarial Machine Learning attacks. In particular, the study investigates both evasion and inference attack scenarios targeting supervised learning models operating on Controller Area Network (CAN) traffic.

Dataset Pre-Processing. The OTIDS dataset [6], containing Normal, DoS, Fuzzy, and Impersonation CAN traffic, was adopted for the experiments. Following Bari et al. [7], non-informative features were removed, categorical attributes were converted into numerical values, and class-specific undersampling was applied to reduce dataset size while preserving class distribution. The resulting dataset was standardized through Z-score normalization.

Evasion Attack Setup. The dataset was split into training (A, 60%) and testing subsets (B and C, 20% each), preserving class distributions. Decision Tree (DT), Random Forest bagging based (RF), Gradient Boosting (GB), and were trained as victim IDS models. Hyperparameters were optimized through 5-Fold Stratified Cross Validation on A using the weighted F1-score, while RF was configured with three estimators to reduce adversarial example generation time.

Inference Attack Setup. The dataset was split into A (30%) and B (70%). A was used to train and test the victim models, while B supported the shadow-model phase for generating the Output Shadow Dataset (OSD). The attack model was then evaluated through a binary membership inference task (*in*/*out*). Victim models included DT, RF, GB, and K-Nearest Neighbors (K-NN), whereas attack models were based on DT, RF, GB, LR, and K-NN. Evaluation was performed using the same metrics adopted for the evasion attack.

III. Results

The results of the experiment are provided below, grouped by attack—*Inference Attack* and *Evasion Attack*—reporting comments with respect to accuracy and insights on the selection of the algorithms.

a) Inference Attack: the weighted accuracy of the various attack models hovers around 50%, indicating a failure of the attack (since they perform like a random classifier). From the defense team's perspective, these findings provide valuable insights into the selection of potential ML algorithms for the IDS. Several hypothetical factors may contribute to this failure, including issues with feature extraction and the decision to utilize shallow supervised learning models.

b) Evasion Attack: the results of the "Normal" test, revealing notably high weighted accuracies for DT and RF, both around 99%. In contrast, the weighted accuracies for the other models range from 94% to 96%. The results of the "Adversarial" test, showing that the impact of the ZOO attack on the performance of the ML models is consistent across the board. Only the RF and Gradient Boosting (GB) models exhibit a slightly lower weighted accuracy by 0.01 compared to the others. The effect of the attack is significant, resulting in an approximate 70% decrease in weighted accuracy. The "Defense" test results, suggesting that adversarial training serves as an effective countermeasure against this attack, as evidenced by the increase in weighted accuracy.

IV. Conclusion

The growing integration of Connected and Autonomous Vehicles into Smart City ecosystems improves urban mobility but also introduces new cybersecurity risks. In this work, we evaluated the impact of two Black-Box Adversarial Machine Learning attacks, Zeroth-Order Optimization (ZOO) and Membership Inference (MI), against ML-based IDSs for automotive CAN networks.

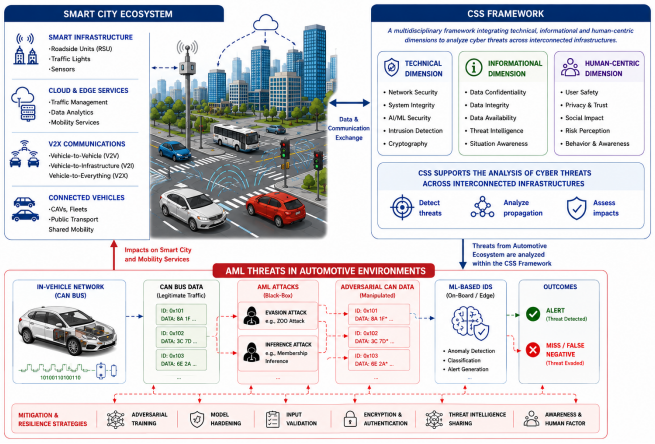

Fig. 1. Cyber-Social Security (CSS) in Smart City Context: Integrating AML Threats in Automotive Environments.

Results showed that ZOO attacks can significantly reduce detection performance, with an average decrease of about 70% in weighted accuracy, while MI attacks proved ineffective under the considered conditions. These findings highlight the potential impact of AML threats not only on vehicle security but also on broader Smart City services, including mobility continuity, public safety, and urban resilience.

The experiments also demonstrated that adversarial training can improve IDS robustness, suggesting that proactive defenses are essential for secure Smart Mobility. Future work will investigate Deep Learning models, V2X environments, and cyber-social impact assessment approaches to support more resilient Smart City transportation systems.

References

[1] D. Garikapati and S. S. Shetiya, "Autonomous vehicles: Evolution of artificial intelligence and the current industry landscape," *Big Data and Cognitive Computing*, vol. 8, no. 4, 2024.

[2] S. Saki and M. Soori, "Artificial intelligence, machine learning and deep learning in advanced transportation systems, a review," *Multimodal Transportation*, vol. 5, no. 1, p. 100242, 2026.

[3] M. Aledhari, R. Razzak, M. Rahouti, A. Yazdinejad, R. M. Parizi, B. Qolomany, M. Guizani, J. Qadir, and A. Al-Fuqaha, "Safeguarding connected autonomous vehicle communication: Protocols, intra- and inter-vehicular attacks and defenses," *Computers & Security*, vol. 151, p. 104352, 2025.

[4] V. Santa Barletta, M. Calvano, and A. Sciacovelli, "Cyber social security in multi-domain operations," in *2024 IEEE International Workshop on Technologies for Defense and Security (TechDefense)*, pp. 41–46, IEEE, 2024.

[5] V. S. Barletta, D. Caivano, C. Catalano, M. de Gemmis, and D. Impedovo, "Cyber social security education," in *International Conference on Extended Reality*, pp. 240–248, Springer, 2024.

[6] H. Lee, S. H. Jeong, and H. K. Kim, "Otids: A novel intrusion detection system for in-vehicle network by using remote frame," in *2017 15th Annual Conference on Privacy, Security and Trust (PST)*, pp. 57–5709, 2017.

[7] B. S. Bari, K. Yelamarthi, and S. Ghafoor, "Intrusion detection in vehicle controller area network (can) bus using machine learning: A comparative performance study," *Sensors*, vol. 23, no. 7, 2023.