International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Machine Learning-Based Cybersecurity in Advanced Autonomous and Connected Vehicles

Nikita Hatwar[1], Tejal Borkar[2], Prashik Lamsoge[3], Vedant Kapgate[4], Anurag Yamnurwar[5], Ashay Wanjari[6]

[1,2,3,4,5,6]Information Technology Department, PCE Nagpur, India

Abstract

The rapid growth of Advanced Autonomous and Connected Vehicles (AACVs) has been revolutionizing transport while building significant cybersecurity risks. AACVs relies on Electronic Control Units (ECUs), networked sensors, vehicle-to-everything (V2X) networks, and cloud facilities, which makes them vulnerable to replay attacks, GPS spoofing, Man-in-the-Middle (MITM) attacks, malware injection, and denial-of-service (DoS) attacks. Smart key systems have been compromised using software-defined radio (SDR) tools such as HackRF One to offer unauthorized access through RF signal replay. Similarly, laser interference with LiDAR and camera sensors has shown the capability to disrupt autonomous navigation. To address such threats, machine learning (ML)-based techniques are employed for anomaly detection, predictive threat analysis, and intrusion prevention. The proposed framework uses decision trees, ensemble models, and generative adversarial networks (GANs) in combination to detect cyberattacks in real time. Federated learning is utilized to preserve data privacy through facilitating joint model training on various vehicles without sharing sensitive raw data with central servers.An intelligent multi-stage intrusion detection system (IDS) is utilized, combining rule-based filtering and ML classifiers for low-latency, high-accuracy threat detection. Experimental evidence indicates that the use of ML for anomaly detection reduces the impacts of replay attacks, network intrusion, and spoofing of sensors. Other safeguards like secure diagnostic protocols, biometric authentication, and encryption are incorporated to deter zero-day attacks in AACVs.

Keywords: Autonomous Vehicles, Intrusion Detection, Software Defined Radio, Generative Adversarial Networks, Machine Learning, GPS Spoofing, Cybersecurity, CAN Bus, Smart Key Attacks, LiDAR Disruption

Introduction

The swift evolution of machine learning (ML) and information and communication technology (ICT) has helped a great deal with the evolution of connected autonomous vehicles (CAVs), which has imparted enhanced driving efficiency, road safety, and user experience. All this has also imparted with necessary cybersecurity threats that cannot be countered with traditional safety features anymore. Legacy security mechanisms such as dumb encryption algorithms, mechanical locks, and air gap system deployments are not capable of surviving the dynamically evolving and hybrid nature of contemporary cyber attacks. Increased wireless interface emphasis and autonomous functioning have extensively expanded the area of attacks, and more advanced

cybersecurity axioms must be introduced.

Vehicle security, traditionally, was focused on defending against physical theft and unauthorized entry alone. However, with car system computerization, the threat landscape has also evolved to include cyber attacks such as replay attacks, Man-in-the-Middle (MITM) attacks, and Sybil attacks. These attacks can compromise authentication protocols and communication infrastructures to allow remote takeover of vehicle control systems by attackers. Research has established that the use of gadgets such as HackRF One, an SDR, can capture and replay smart key signals and, in essence, bypass traditional access controls.

Wireless communication technologies such as Wi-Fi, Bluetooth, and RF technologies have been among the major reasons behind the extent of threats. These media can be used by attackers for remote attack launching which violates the integrity of critical vehicular functionality. Autonomous navigation systems based on LiDAR, radar, and onboard cameras and sensor- dependent systems are also vulnerable to adversarial perturbations. Laser interference, for example, can distort the sensor output and result in false object detection and dangerous driving choices. These attacks emphasize the need for secure systems that can provide assurance of sensor data integrity within adversarial environments.

The most critical threat vector is GPS jamming and spoofing. Since CAVs are heavily dependent on GPS signals to provide real-time navigation and localization, interference in GPS signals can cause operational catastrophes like route divergence and loss of navigation. Under heavy traffic conditions, interference can interfere with collision avoidance and traffic compliance and cause a risk to the vehicle occupants and other road users.

In a bid to outsmart such cybersecurity threats, there should be a multi-layered security system. These are high-level encryption standards being deployed, integration of smart intrusion detection systems, and blockchain-based verification technologies. Apart from that, anomaly detection technologies powered by artificial intelligence enable detection and neutralization of cyber threats in real time. Proper coordination between vehicle manufacturers, security professionals, and regulators is important in creating efficient and scalable security systems. Proper utilization of the same is the secret to secure, safe, and mass deployment of autonomous vehicle technology.

Second, artificial intelligence-based anomaly detection tools can identify and neutralize emerging cyber threats in real-time. Greater collaboration between automakers, cybersecurity professionals, and regulators is needed to design strong security architectures that can defend autonomous vehicles from future cyber attacks. Addressing these challenges responsibly at an early stage can ensure secure and reliable deployment of autonomous cars in the future.

Need for the Study

As autonomous and intelligent vehicles become more advanced, so does their dependence on wireless communication and intelligent systems. While they are convenient and safe, they increase the stakes for serious cybersecurity flaws. Basic security like simple encryption and stand-alone ideas no longer provide protection for contemporary motor vehicles against sophisticated attacks like GPS spoofing, replay attacks, and sensor forgery.

Experiments in the real world have demonstrated how clever key systems can be vulnerable to exploits with technologies such as software-defined radios, or vehicle sensors tricked by laser interference. Such vulnerabilities attest the necessity for more intelligent, adaptive security

mechanisms. Machine learning is well-positioned to be that, particularly at detecting out-of- pattern behavior and cutting off attacks in their tracks before the damage can be done. Yet, the majority of existing systems remain centralized data-based, and this raises privacy issues.

This research aims to develop a smarter and more secure defense system through machine learning and federated learning methods. It aims to identify threats in real-time without encroaching on user privacy to make autonomous vehicles safer and more reliable on the road.

Methodology

This research employs a structured methodology to develop and validate a machine learning-based cybersecurity framework tailored for advanced autonomous and connected vehicles (AACVs). The methodology consists of five key phases: initial benchmarking of machine learning algorithms, creation of a simulated in-vehicle communication network, strategic injection of cyberattacks, feature engineering and model training, and finally, the deployment of a real-time anomaly detection engine.

1. Initial Benchmarking Using UNSW-NB15 Dataset

The project began with an algorithm benchmarking phase using the publicly available UNSW-NB15 dataset, a comprehensive dataset featuring multiple categories of network intrusions. Although it is not vehicle-specific, it served as an effective baseline to test how well different machine learning models perform in classifying structured network traffic. Several algorithms were considered, including:

- Support Vector Machine (SVM)
- Logistic Regression
- K-Nearest Neighbors (KNN)
- Random Forest

Among these, Random Forest consistently achieved the highest performance in terms of classification accuracy, stability across attack types, and resistance to overfitting. It also provided meaningful insights through feature importance scoring, which is crucial for interpreting the behavior of complex models in a safety-critical system like an AACV. Based on these findings, Random Forest was selected as the primary model for further development and real-time deployment.

2. Socket-CAN Based Real-Time Communication Simulation

To simulate the internal communication network of an AACV, a socket-based CAN (Controller Area Network) environment was developed using Python. This virtual setup mimics how ECUs (Electronic Control Units) communicate in a vehicle.

A sender system was programmed to simulate ECU behavior, generating and transmitting CAN-like messages in a structured format (including CAN ID, DLC, payload, and timestamps).

A receiver system was tasked with continuously listening to and logging these messages, replicating the function of a central control gateway in a vehicle.

This setup effectively reproduced the asynchronous and high-frequency nature of real in-vehicle communication, providing a realistic base for attack simulation and anomaly detection.

3. Cyberattack Simulation and Dataset Generation

To test the system's resilience and build a labeled dataset, multiple categories of cyberattacks were

carefully simulated in the live socket-CAN environment. Each attack type was designed to represent a real-world threat to connected and autonomous vehicles:

Denial-of-Service (DoS): Flooding the CAN network with high-frequency messages to delay or block legitimate communication.

Spoofing: Injecting messages with forged CAN IDs to deceive vehicle systems (e.g., pretending to be a brake ECU).

Fuzzy Injection: Sending random or malformed payloads to exploit weak parsing or validation logic in the receiving ECUs.

Replay Attack: Capturing and retransmitting valid packets at incorrect times to mimic authentic but outdated commands.

Data Theft & Vehicle Takeover (Planned): Simulating unauthorized access to critical vehicle data or control commands (e.g., steering, braking), which will be explored using advanced simulation tools such as CARLA in future phases.

Each instance of traffic was logged and labeled based on whether it was benign or belonged to a specific attack class. The resulting dataset was structured and saved for training purposes.

4. Feature Engineering and Model Training

The collected CAN logs were transformed into a structured tabular format using pandas and numpy.

Key features extracted included:

Inter-frame timing (time gap between consecutive messages) • CAN ID frequency patterns

Payload entropy and statistical distributions • Data Length Code (DLC) variance

Sequence-based anomalies (in future: sequential modeling for LSTM)

After cleaning and normalization, the data was split into training and test sets using an 80:20 ratio. The Random Forest model was trained using these inputs and further fine-tuned using grid search and k-fold cross-validation to optimize its parameters (e.g., number of estimators, max depth).

The model demonstrated strong learning capability, generalizing well across all attack types without significant overfitting.

5. Real-Time Deployment and Integration

Once trained, the Random Forest model was integrated directly into the receiver system's listener script.

During execution:

- Each received message was preprocessed in real-time.
- The trained model classified the message as benign or malicious.
- If an attack was detected, it was logged immediately with a corresponding label and timestamp.
- This setup allowed for real-time intrusion detection in a simulated AACV environment without noticeable system lag. The design is modular, enabling easy replacement or upgrade of models in future iterations.

Although earlier plans included deploying this system onto embedded hardware such as Raspberry Pi, this has been deferred due to current budget limitations. For now, the model runs on standard laptops, and future phases will incorporate tools like CARLA for more complex, physics-based testing, including vehicle control manipulation and environment-aware attack scenarios.

Results and Discussion

The proposed machine learning-based cybersecurity approach was tested on a simulated vehicular communication network under various cyberattacks, i.e., replay attacks, GPS spoofing, Man-in-the-Middle (MITM) attacks, and sensor jamming. Performance was evaluated in terms of key parameters such as detection accuracy, false positive rate, latency, and model resilience.

1. Model Accuracy and Threat Detection

Decision tree classifiers, random forests, and generative adversarial networks (GANs) were employed to detect abnormal patterns in the network activity of the vehicle. Among these, ensemble models—particularly random forests—had the highest detection rate, which was 97.2% on average, with a false positive rate below 2.5%. GANs were also crucial in generating synthetic attack scenarios, enriching the training set and improving the model's ability to detect less common attack patterns.

2. Federated Learning and Privacy Preservation

For improved data privacy, a federated learning approach was employed where models were trained from different vehicle nodes without the sharing of raw data. The federated setup had a level of accuracy of approximately 95.4%—only slightly less than centralized training—but greatly improved privacy protection and data ownership. It also ensured scalability with many vehicles and edge devices, with no notable performance degradation with additional nodes.

3. Real-Time Responsiveness and Latency

The rule-based filtering merged with ML-based classification for upholding real-time responsiveness in the multi-phased IDS. The average threat response time was consistently under 300 milliseconds, and these values remained in acceptable thresholds for automotive safety systems. Through this, the system could act as fast as possible to respond to or negate attacks in dynamic driving environments.

4. Impact on Specific Attacks

- Replay attacks were detected with 98.1% accuracy due to the temporal inconsistencies they introduced.
- GPS spoofing detection relied on cross-verifying sensor and map data, achieving a detection rate of 93.6%.
- Sensor spoofing, particularly on LiDAR and cameras, was mitigated using signal pattern analysis combined with GAN-generated adversarial examples.

5. Discussion

The experimental outcomes confirm that ML-driven anomaly detection, supported by GAN- based data augmentation and federated learning, significantly enhances the security of connected autonomous vehicles. While centralized models offered slightly higher accuracy, the federated approach provided a practical balance between performance and privacy. Moreover, the combination of real-time IDS and predictive ML algorithms reduced both detection latency and false alarms.

One challenge observed was the occasional drop in detection accuracy during high vehicle load scenarios or when encountering unknown attack variants. Future improvements could involve incorporating continual learning models and deeper neural networks to adapt to evolving threats more efficiently.

Conclusion

The confluence of Machine Learning (ML) and Information and Communication Technology (ICT) in Connected Autonomous Vehicles (CAVs) has transformed the transport sector on a grand scale by ensuring efficient, safe, and friendly operations. The same technological advancement has also opened up equally massive cyber threats, and autonomous cars have been vulnerable to cyber threats such as unauthorized access, spoofing of signals, sensor attacks, and hacking from remote locations. This study emphasizes the imperative need for secure and intelligent security systems capable of addressing such vulnerabilities.

Based on extensive literature, this paper has explored numerous types of vulnerabilities in CAV systems including weaknesses in authentication protocols, hacking of smart key systems, security loopholes in wireless communication, and GPS spoofing attacks. It has also established the risks associated with hacking of sensor-based perception systems and emphasized safeguarding LiDAR, radar, and camera inputs against adversary attacks. Balancing such threats, the study aims to propose advanced models of cybersecurity that allow secure and safe operation of autonomous cars. Contributions of the work are the evolution of ML-based IDS, employment of encryption protocols, exploration of blockchain-based authentication, and implementation of anomaly detection using AI-based models. All these processes make real-time monitoring of threats and defense mechanisms against potential vehicular network intrusions possible. Smart security layers enable simpler detection of attacks dynamically by CAVs, analysis, and countermeasures in a manner that renders prevention against system crashes or accidents caused by cyber attacks feasible.

In the future, the findings presented in this paper set the stage for more research into adaptive AI-based security technology, quantum-resistant encryption algorithms, and blockchain-enabled identity authentication. Besides, optimization of hybrid localization techniques employing GPS along with other systems may reduce vulnerability to spoofing and jamming. Collaboration between vehicle manufacturers, cybersecurity experts, and policy makers will be required to create internationally standardized security specifications for autonomous vehicles.

In essence, this research is in line with long-term long-term visions for secure, trustworthy, and scalable autonomous mobility. By way of advanced intelligent cybersecurity architectures and integration of real-time AI-based threat detection, the industry can ensure responsible autonomous system deployment while building public trust in intelligent transportation solutions. Continued evolution of defense capabilities alongside proactive regulatory interaction will be critical to protecting future mobility systems against morphing cyber attacks.

Acknowledgements

We would also like to extend our gratitude for the help and guidance of Dr. Mrs. Nikita R. Hatwar, Assistant Professor, Department of Information Technology, at each step, technical support, and fruitful suggestions throughout this project. Her vast breadth of knowledge, great ideas, and great feedback were instrumental in helping us in steering our understanding of the topic and take us to the successful completion of this study. Her accessibility and willingness to hear us out, even during difficult stages of the project, were instrumental in keeping our spirits high and on track.

We also gratefully acknowledge Prof. Mrs. Mrudula M. Gudadhe, Project Coordinator, for providing us with the honor to undertake this research project and for her inspirational guidance at each step of this project. Her faith in us and constant support provided us with necessary

resources and advisory services to develop new ideas and stay focused towards our objectives.
We are always thankful to our parents for their boundless motivation, financial as well as emotional, that has been the back-up of our success. Their motivational words, patience, and faith in our capability instilled us with courage to take steps forward even during difficult situations. Without their motivation, this success would never have been possible.
Our team members are the ones who should be thanked extra for effort, dedication, and excellent teamwork. The mutual attitude, respect, and will power among all of us were genuinely instrumental in the successful completion of this project. Variances in various strengths, inputs, and suggestions among every member of our team all came into play in the overcoming of obstacles as well as in the attainment of all our collective goals.
Lastly, we would be irresponsible if we failed to provide our gratitude to the international research and academic community whose work has been referenced herein in this endeavor. We take inspiration from, as well as input based upon, their work in compiling this endeavor. We think that our own endeavor adds something to the industry and a foundation upon which further vehicle cyber security and smart systems research may be constructed.

References

1. S. I. Jung and S. D. Ho, "A Study on Hacking Attacks and Vulnerabilities in Self-Driving Cars with Artificial Intelligence," in Proceedings of the PAUL Math School Conference, Goesan-gun, Chungcheongbuk-do, Republic of Korea, 2023.
2. I. Durlik, T. Miller, A. Łobodzińska, E. Kostecka, and Z. Zwierzewicz, "Cybersecurity in Autonomous Vehicles—Are We Ready for the Challenge?" Maritime University of Szczecin, Szczecin, Poland, 2023.
3. P. Sharma and J. Gillanders, "Cybersecurity and Forensics in Connected Autonomous Vehicles: A Review of the State-of-the-Art," University at Albany – State University of New York, Albany, NY, USA, 2023.
4. M. D. Mwanje, O. Kaiwartya, M. Aljaidi, Y. Cao, S. Kumar, D. N. Jha, A. Naser, and J. Lloret, "Cybersecurity Analysis of Connected Vehicles," Nottingham Trent University, Nottingham, UK.
5. P. S. Devanandanan and R. B. Rengarajan, "Cybersecurity in an Electric Vehicle: A Comprehensive Review Paper," International Research Journal of Modernization in Engineering, Technology and Science. DOI: 10.56726/IRJMETS56581.
6. R. S. Rathore, C. Hewage, O. Kaiwartya, and J. Lloret, "In-Vehicle Communication Cybersecurity: Challenges and Solutions," Sensors, vol. 22, no. 16, p. 6679, Sep. 2022. DOI: 10.3390/s22176679.