Sustainable Intrusion Detection in Vehicular Controller Area Networks using Machine Intelligence Paradigm

Ahmed A. Metwaly1, and Ibrahim Elhenawy2, *

1,2 Faculty of Computers and Informatics, Zagazig University, Zagazig, Sharqiyah 44519, Egypt; a.metwaly23@fci.zu.edu.eg; ielhenawy@zu.edu.eg.

* Correspondence: ielhenawy@zu.edu.eg.

Abstract: The advent of smart mobility and the proliferation of connected vehicles have introduced new challenges in securing Vehicular Controller Area Networks (CANs) against cyber threats. This paper proposes an innovative machine intelligence paradigm for sustainable intrusion detection within vehicular networks. We present a Deep Neural Network (DNN) model that effectively classifies CAN traffic into categories, including Normal, Denial of Service (DoS), Gear Attack (Spoofing), RPM Attack (Spoofing), and Fuzzy Attack. The DNN’s architecture is designed to learn and adapt to the dynamic nature of vehicular communications, enhancing its ability to detect network intrusions. The study encompasses an inclusive exploration of the CAN bus architecture, message data format, and related security vulnerabilities to provide a solid foundation for intrusion detection. Our methodology employs mathematical representations of the DNN model, offering insight into its training process. Visualizations of results, such as confusion matrices, ROC-AUC curves, T-SNE plots, and SHAP explanations, provide a holistic view of the model’s performance and offer valuable insights for system refinement. By bridging the gap between machine intelligence and vehicular security, this research contributes to the ongoing efforts to fortify critical infrastructure, ensuring the reliability and sustainability of vehicular networks in the era of connected and autonomous vehicles.

Keywords: Machine Intelligence, Intrusion Detection, Vehicular Controller Area Networks, Deep Neural Network, Cybersecurity, CAN Bus, Network Security, Smart Mobility, Automotive Technology, Threat Classification.

1. Introduction

In the era of rapidly advancing technology, the integration of smart vehicles into our daily lives has revolutionized the way we commute, communicate, and experience transportation. Vehicular Controller Area Networks (CANs) lie at the heart of this transformation, serving as the nervous system that enables seamless communication and coordination among various electronic control units within modern vehicles. However, as the automotive industry embraces connectivity and automation, the susceptibility of these networks to cyber threats becomes an increasingly critical concern [1-2].
The automotive ecosystem, once confined to mechanical engineering, has now evolved into a sophisticated fusion of hardware and software components. While this evolution has brought unprecedented convenience and efficiency, it has also exposed vehicles to a new realm of cybersecurity challenges [4]. The proliferation of smart features, Internet of Things (IoT) integration, and the prospect of autonomous driving have expanded the attack surface, making vehicular networks a prime target for malicious actors seeking to compromise safety, privacy, and functionality [3-5].

In response to these emerging threats, traditional security mechanisms have proven inadequate. The conventional firewall and signature-based intrusion detection systems (IDS) struggle to keep pace with the evolving tactics and techniques of cyber adversaries. To address this gap, an innovative approach is required, one that not only offers robust security but also aligns with the broader goals of sustainability and environmental responsibility [6-7]. This paper introduces a novel paradigm for intrusion detection in Vehicular CAN—leveraging the power of Machine Intelligence to create a sustainable and adaptive defense mechanism. By combining cutting-edge machine learning and artificial intelligence techniques with a deep understanding of automotive systems, we propose a holistic framework capable of identifying and mitigating intrusions while minimizing false positives and reducing energy consumption [8-10].

Our paper is structured as follows: In Section 2, we provide an overview of the existing research in vehicular network security. Section 3 outlines our innovative machine intelligence framework for intrusion detection. In Section 4, we describe the setup and key parameters used in our experiments. Section 5 presents a thorough analysis of the experimental outcomes and engages in a comprehensive discussion. Finally, in Section 6, we summarize our findings and outline future research directions. Table 1 outlines the content of each section.

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Related Work</td>
<td>Review of related literature and research</td>
</tr>
<tr>
<td>3. Methodology</td>
<td>Presentation of our machine intelligence framework including algorithms, techniques, and implementation details</td>
</tr>
<tr>
<td>4. Experimental Configurations</td>
<td>Details on the experimental setup, datasets, and parameters</td>
</tr>
<tr>
<td>5. Results and Discussion</td>
<td>In-depth analysis of experimental results, including quantitative and qualitative insights</td>
</tr>
<tr>
<td>6. Conclusion</td>
<td>Summary of key findings, contributions, and future research directions</td>
</tr>
</tbody>
</table>
In this section, we embark on a journey through the existing body of research, seeking to understand the current state-of-the-art in intrusion detection for Vehicular CANs. By examining the work of pioneers and innovators in the field, we aim to identify key challenges, trends, and insights that lay the foundation for our proposed sustainable intrusion detection framework. Hossain et al. [11] introduced a Long Short-Term Memory-Based Intrusion Detection System for in-Vehicle Controller Area Network Bus. Their work is significant as it addresses the critical issue of security within the context of in-vehicle communication networks. By implementing Long Short-Term Memory (LSTM) networks, they provided a promising approach to identifying and preventing intrusions within CANs. This research contributes to the field by presenting an effective method to enhance the security of in-vehicle networks, which is of paramount importance in the era of connected and autonomous vehicles. Mehedi et al. [12] presented a Deep Transfer Learning Based Intrusion Detection System for Electric Vehicular Networks. Their approach, based on deep transfer learning, is of great significance due to the growing prevalence of electric vehicles. The study offers a solution to the unique security challenges posed by these vehicles, emphasizing the adaptability and transferability of deep learning models. Tomlinson et al. [13] focused on advancements in intrusion detection methods for the Automotive Controller Area Network. Their work is vital as it recognizes the need for robust security in automotive systems. By exploring the viability of intrusion detection solutions, they contribute to the development of security measures in automotive communication networks, safeguarding the reliability and safety of modern vehicles. Sharmin and Mansor [14] explored the application of machine learning for Intrusion Detection on the In-Vehicle Network. Their research was relevant in light of the increasing integration of machine learning techniques into network security. By applying machine learning to in-vehicle networks, the study highlighted the potential for more dynamic and adaptable intrusion detection systems, offering a valuable contribution to the evolving landscape of network security. Shahriar et al. [15] introduced CANShield, a Deep Learning-Based Intrusion Detection Framework for CANs at the Signal-Level. This approach is noteworthy as it operates at the signal level, offering fine-grained security. The utilization of deep learning methods further enhances the precision of intrusion detection in CANs, making this research an important contribution to ensuring the integrity of automotive communication systems. Olufowobi et al. [16] developed Saiducant, which is a Specification-Based Automotive Intrusion Detection System Using Controller Area Network (CAN) Timing. Their work was significant as it introduces a specification-based approach to intrusion detection. By considering the timing aspects of CAN messages, their study offered a method that complements existing security measures, enhancing the overall security posture of automotive networks. Nam et al. [17] proposed an Intrusion Detection Method Using Bi-Directional GPT for in-Vehicle CANs. By leveraging the capabilities of GPT-based models, this research opens new avenues in natural language processing and machine learning for intrusion detection. The approach is innovative in its bi-directional use of GPT models, making it a valuable contribution to in-vehicle network security. Alfardus and Rawat [18] presented an Intrusion Detection System for CAN Bus In-Vehicle Network Based on Machine Learning Algorithms. Their work is crucial in the context of in-vehicle network security, as it emphasizes the utilization of machine learning algorithms to safeguard CANs.
This study underscores the adaptability and efficiency of machine learning in intrusion detection, making it a relevant addition to the field. Islam et al. [19] introduced a Graph-Based Intrusion Detection System for CANs. Their research offered a unique approach by using graph-based techniques for intrusion detection. By modeling the network as a graph and analyzing its structure, their study contributes to the development of novel methods for enhancing the security of CANs.

3. Methodology

In this section, we delve into the core of our research, unveiling the intricacies of the methodology that underpins our proposed sustainable intrusion detection framework for Vehicular Controller Area Networks (CANs) using a machine intelligence paradigm.

3.1. System Architecture

The Controller Area Network (CAN) bus architecture serves as the foundational communication framework within Vehicular Controller Area Networks. It’s characterized by a robust and distributed structure where multiple electronic control units (ECUs) interconnect through a shared bus. This architecture enables real-time, low-latency communication among ECUs, facilitating critical functions such as engine control, safety systems, and infotainment. The CAN bus relies on a two-wire differential pair to transmit messages, with a dominant “0” and a recessive “1” encoding scheme. Messages are broadcast to all ECUs on the network, and each ECU filters messages based on their unique identifiers. However, this open and decentralized nature of the CAN bus also exposes it to potential security vulnerabilities, including message spoofing, unauthorized access, and denial-of-service (DoS) attacks. Understanding the intricacies of the CAN bus architecture,
its message format, and associated security risks is fundamental to the development of an effective intrusion detection system within vehicular networks.

Messages within the CAN bus are structured with specific components, including an identifier, data bytes, and control bits. The identifier, often referred to as the CAN ID, specifies the message’s priority and source ECU. Data bytes contain the actual information being transmitted, ranging from 0 to 8 bytes in length, with each byte encoding critical vehicle data. Control bits include elements like the identifier extension bit (IDE) and the remote transmission request (RTR) bit, which affect message transmission and interpretation. This standardized data format ensures uniform communication across the network. However, it also introduces potential vulnerabilities, as attackers can manipulate message content or impersonate legitimate ECUs to send malicious messages, compromising vehicular network security (See Figure 1).

The CAN bus, while efficient for real-time communication, faces several security vulnerabilities that demand attention in the context of intrusion detection. Message spoofing, where attackers impersonate trusted ECUs by manipulating CAN IDs, can lead to unauthorized control of critical vehicle functions. Additionally, unauthorized access may occur when malicious actors gain entry to the network, potentially compromising its integrity. Denial-of-service (DoS) attacks pose another threat, where attackers flood the bus with excessive messages, disrupting normal communication and causing potential safety hazards. The open nature of the CAN bus architecture further exacerbates these vulnerabilities, necessitating the implementation of robust intrusion detection mechanisms to safeguard vehicular networks against cyber threats.

3.2. Machine Intelligence Model

The process of classifying CAN traffic into distinct attack categories involves the construction of a Deep Neural Network (DNN) tailored to the unique characteristics of vehicular networks. A DNN is a powerful machine learning architecture consisting of multiple interconnected layers of artificial neurons. It excels at learning complex patterns and relationships within data, making it well-suited for intrusion detection tasks. Mathematically, the DNN can be represented as follows:

Let X represent the input data, which consists of CAN traffic attributes, and Y represent the output, which corresponds to the attack categories (Normal, DoS, Gear Attack, RPM Attack, Fuzzy Attack). The DNN comprises multiple hidden layers (commonly referred to as H) connected through weighted connections (represented as W) and activated by nonlinear functions (often denoted as σ). The mathematical representation of a neuron in a hidden layer can be expressed as follows:

\[
Z^{[h]} = W^{[h]}X + b^{[h]}
\]

\[
A^{[h]} = \sigma(Z^{[h]})
\]

(1)

where \(Z^{[h]}\) is the weighted sum of inputs in layer \(h\), \(A^{[h]}\) is the activation output of the neurons in layer \(h\), \(W^{[h]}\) represents the weight matrix for layer \(h\), \(b^{[h]}\) is the bias term for layer \(h\), \(\sigma\) denotes the activation function (e.g., ReLU, Sigmoid) applied element-wise to \(Z^{[h]}\).
The DNN is trained using a labeled dataset that associates input CAN traffic samples (X) with their corresponding attack categories (Y). The training process aims to optimize the weight and bias parameters to minimize a loss function (often represented as L) that quantifies the difference between predicted and actual attack categories:

$$L = -\frac{1}{n} \sum_{k=1}^{n} (t_i \log y_i + (1 - t_i) \log (1 - y_i)),$$

(2)

Where $y_i$ represents the predicted attack categories produced by the DNN. Through an iterative process known as backpropagation, the DNN adjusts its parameters using optimization techniques (e.g., stochastic gradient descent) to minimize the loss function. Once trained, the DNN can classify incoming CAN traffic into one of the predefined categories, effectively detecting and categorizing network intrusions.

4. Experimental Design

In this section, we embark on a journey through the practical implementation of our sustainable intrusion detection framework in Vehicular CANs. Having elucidated the methodology in the previous section, we now shift our focus to the real-world deployment and assessment of our innovative system. We detail the experimental setups, datasets, and configurations that were meticulously designed to evaluate the performance, robustness, and sustainability of our intrusion detection solution. For the execution of our experiments, a dedicated and meticulously designed implementation setup was employed. The hardware specifications included a high-performance server equipped with a multi-core processor (Intel Core i7, 3.5 GHz), ample RAM (32 GB DDR4), and solid-state drives (SSDs) for rapid data access. This robust hardware infrastructure ensured the computational capacity required for the intricate tasks of intrusion detection and analysis. The server ran on a Linux-based operating system, Ubuntu 20.04 LTS, providing a stable and well-supported platform for our experiments. To harness the power of machine learning and deep learning techniques, we employed open-source libraries and frameworks, including TensorFlow, and Scikit-learn, for model development and evaluation. Additionally, for data preprocessing, visualization, and analysis, we utilized Python programming with Pandas, Matplotlib, and Seaborn. This carefully orchestrated hardware and software ensemble formed the foundation of our experimental environment, enabling us to rigorously evaluate the performance of our sustainable intrusion detection system in Vehicular CAN. Our study employs a Car-Hacking dataset, the distribution of which is meticulously outlined in Table 2. This dataset serves as the foundation of our experimental endeavors, enabling us to delve into the intricate realm of car security and vulnerability analysis. With the comprehensive insights provided by Table 2, we are well-equipped to investigate and address the critical challenges and threats associated with automotive cybersecurity [20].

### Table 2: Summary of Attributes of Car-Hacking Dataset.

<table>
<thead>
<tr>
<th>Attack Type</th>
<th>Normal</th>
<th>Denial of Service (DoS)</th>
<th>Gear Attack (Spoofing)</th>
<th>RPM Attack (Spoofing)</th>
<th>Fuzzy Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timestamp Interval</td>
<td>N/A</td>
<td>0.3 milliseconds</td>
<td>1 millisecond</td>
<td>1 millisecond</td>
<td>0.5 milliseconds</td>
</tr>
<tr>
<td>CAN ID Pattern</td>
<td>Random Hexadecimal</td>
<td>Dominant “0000”</td>
<td>Random Hexadecimal</td>
<td>Random Hexadecimal</td>
<td>Random Hexadecimal</td>
</tr>
</tbody>
</table>
Data Byte Range | 0–8 (all 8 bytes) | 0–8 (all 8 bytes) | 0–8 (all 8 bytes) | 0–8 (all 8 bytes) | 0–8 (all 8 bytes)
--- | --- | --- | --- | --- | ---
Data Pattern | Varied | Varied | Varied | Varied | Varied
Sample Class | Legitimate | Malicious | Malicious | Malicious | Malicious
Number of Samples | 988872 | 587521 | 597252 | 654897 | 491847

5. Results Discussion

Having laid the foundation of our methodology and detailed the experimental configurations, we now embark on the crucial phase of our research journey—presenting the results and engaging in an insightful discussion. In this section, we unveil the outcomes of our experiments, showcasing the performance and efficacy of our sustainable intrusion detection framework within Vehicular CANs. Through a meticulous analysis of these results, we aim to shed light on the strengths, limitations, and implications of our approach.

Figure 2 presents a visual representation of our model’s performance through a confusion matrix. This matrix provides a comprehensive snapshot of the classification results, categorizing instances into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The diagonal elements (TP and TN) represent correctly classified instances, while off-diagonal elements (FP and FN) indicate misclassifications. This visualization offers critical insights into the effectiveness of our model. We observe a substantial number of true positives, demonstrating the system’s capability to accurately identify malicious intrusions, a fundamental objective in enhancing vehicular network security. Conversely, the occurrence of false positives implies instances where benign activities were incorrectly flagged as malicious. This aspect warrants further investigation to fine-tune the model’s threshold for alarm triggering. The presence of false negatives indicates instances where actual attacks were missed by the system, signaling the need for model enhancements to bolster its sensitivity.

Figure 3 showcases the Receiver Operating Characteristic (ROC) curve along with the Area Under the Curve (AUC) metric, offering a comprehensive evaluation of our
intrusion detection model’s performance. The ROC curve visually illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) across varying classification thresholds. A well-defined curve, rising sharply toward the upper-left corner, signifies a model with superior discriminatory power. The AUC quantifies the model’s ability to distinguish between benign and malicious instances, with a higher AUC indicating better discrimination. In our ROC-AUC curve, we observe a graceful curve that significantly extends towards the top-left corner, indicative of our model’s ability to effectively separate normal network activity from intrusion attempts. Moreover, the AUC metric registers a commendable value, reinforcing the model’s strong discriminatory capabilities. These results underscore the model’s potential for robust intrusion detection within Vehicular Controller Area Networks. However, further analysis of the ROC curve suggests potential areas for fine-tuning the model’s operating threshold to optimize its performance. This visual assessment not only validates the efficacy of our approach but also provides actionable insights for potential enhancements, ensuring the continued security and resilience of vehicular networks.
In Figure 4, we present the learning curves that offer valuable insights into the performance and training dynamics of our intrusion detection model. Learning curves display how key performance metrics, namely accuracy, and loss, evolve as the model undergoes training over multiple epochs. The training curve represents the model’s performance on the training dataset, while the validation curve reflects its performance on a separate validation dataset. A close examination of these curves provides a nuanced understanding of the model’s learning process. Initially, during the early epochs, we observe a convergence of the training and validation curves, indicating that the model rapidly learns to generalize from the training data. However, as training progresses, a divergence may emerge, where the training curve continues to improve while the validation curve stabilizes or slightly regresses. This divergence suggests potential overfitting, prompting us to consider regularization techniques or adjustments to the model architecture. Conversely, if both curves continue to improve without significant divergence, it reflects the model’s ability to generalize effectively.

In Figure 5, we present a T-distributed Stochastic Neighbor Embedding (T-SNE) plot as an insightful visualization of our model’s predictions. T-SNE is a powerful dimensionality reduction technique that allows us to visualize high-dimensional data in a two-dimensional space while preserving the inherent structure and relationships among data points. Each point on the plot represents a data instance, color-coded based on the model’s prediction—benign or malicious. The T-SNE plot reveals clusters and patterns within the data, offering valuable insights into how well our intrusion detection system discerns between different classes of network traffic. Clusters of points closely aligned with one another indicate that the model has successfully grouped similar instances together, reflecting its ability to distinguish between benign and malicious activities. Conversely, points that are more scattered suggest potential areas of improvement where the model’s predictions may be less consistent. This visualization serves as a powerful diagnostic tool, helping us gain a deeper understanding of our model’s performance beyond traditional accuracy metrics. It enables us to identify potential areas for refinement and offers a holistic view of the model’s effectiveness in securing Vehicular Controller Area Networks.

Figure 6 offers a compelling insight into our model’s predictive decisions through SHAP (SHapley Additive exPlanations) explanations. SHAP values provide a comprehensive view of feature importance, shedding light on the factors driving individual predictions. Each point on the plot corresponds to a specific prediction instance, and the horizontal position of the points indicates the SHAP value’s magnitude, with positive values signifying features that push the prediction towards the ‘malicious’ class and negative values indicating features favoring the ‘benign’ class. By visualizing these SHAP explanations, we gain a deeper understanding of the model’s decision-making process. Notably, instances where SHAP values exhibit significant deviations from zero signify strong contributing factors, offering interpretability into why the model classified a given instance as benign or malicious. This visualization not only enhances model transparency but also empowers us to identify and validate the critical features in intrusion detection, thereby strengthening our ability to secure Vehicular CAN.
6. Conclusions

This paper presents a groundbreaking approach to bolster the security of Vehicular Controller Area Networks (CANs) through the innovative integration of machine intelligence. Our research has showcased the development of a deep neural network (DNN) model that effectively classifies CAN traffic into multiple categories, including Normal, Denial of Service (DoS), Gear Attack (Spoofing), RPM Attack (Spoofing), and Fuzzy Attack. By leveraging the power of machine learning, we have demonstrated the model’s robustness in detecting and categorizing network intrusions, thus enhancing the resilience of vehicular networks in the face of evolving cyber threats. Additionally, through the visualization of results, including confusion matrices, ROC-AUC curves, T-SNE plots, and SHAP explanations, we have provided valuable insights into the model’s performance, strengths, and areas for improvement.

The implications of our research extend beyond vehicular networks to broader applications in cybersecurity. This work contributes to the ongoing discourse on safeguarding critical infrastructure by harnessing the potential of machine intelligence. As we look toward the future of smart mobility, the development and deployment of advanced intrusion detection systems will be instrumental in ensuring the security, reliability, and sustainability of vehicular networks. We envision further research and refinement of our approach to continue strengthening the defense mechanisms against cyber threats in the ever-evolving landscape of automotive technology.

Supplementary Materials

Not applicable.

Author Contributions

All Authors contributed equally to this work.

Funding

This research was conducted without external funding support.

Ethical approval
This article does not contain any studies with human participants or animals performed by any of the authors.

**Conflicts of Interest**

The authors declare that there is no conflict of interest in the research.

**Institutional Review Board Statement**

Not applicable.

**Informed Consent Statement**

Not applicable.

**Data Availability Statement**

All data supporting the findings of this study are included within the paper.

**References**


