

Analyzing Car-Hacking Dataset to Understand Cyberattacks and Intrusion Detection in Autonomous Vehicles

Isabela Regina Aguilar, Cristiane Neri Nobre, Henrique Cota de Freitas

Department of Computer Science

Pontifical Catholic University of Minas Gerais (PUC Minas)

Av. Dom José Gaspar, 500, Belo Horizonte, MG, 30535-901, Brazil

iraguilar@sga.pucminas.br, {nobre, cota}@pucminas.br

ABSTRACT

As autonomous and connected vehicles become increasingly popular, the security of in-vehicle networks has emerged as a critical challenge for operations research (OR), requiring systematic optimization approaches. This study applies OR principles to cybersecurity defense systems by analyzing intrusion detection as a multi-criteria decision optimization problem. We evaluate the Car-Hacking dataset from a decision-theoretic perspective, quantifying its effectiveness in developing security countermeasures across multiple attack vectors in automotive Controller Area Network (CAN) environments. Our methodological framework implements multiple algorithmic approaches, Decision Tree, Random Forest, Gradient Boosting, and Convolutional Neural Networks, as competing decision optimization models for security classification tasks. The comparative evaluation demonstrates how OR techniques can effectively balance the trade-off between detection sensitivity and specificity when optimizing in-vehicle security systems. This work contributes to the OR literature by establishing a quantitative framework for evaluating security datasets and demonstrating how structured decision spaces enhance the development of adaptive defense mechanisms for critical transportation infrastructures. The results highlight the value of representative datasets in building robust security optimization models capable of addressing emerging threats in increasingly complex vehicular networks.

KEYWORDS. Operations Research. Intrusion Detection Systems. Vehicular Networks Security.

Paper topics:

D&SP - OR in Defense and Security. AI - OR and AI. AS&DS - Analysis and Data Science.

The authors would like to thank the National Council for Scientific and Technological Development of Brazil (CNPq - Codes: 311573/2022-3 and 311697/2022-4), the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES - Grant PROAP 88887.842889/2023-00 - PUC/MG, Grant PDPG 88887.708960/2022-00 - PUC/MG - Informatics and Finance Code 001), the Foundation for Research Support of Minas Gerais State (FAPEMIG - Codes: APQ-03104-24, APQ-03076-18, APQ-05058-23 and PCE-00344-25).

1. Introduction

Modern vehicles have evolved from simple means of transportation to Intelligent Transportation Systems (ITS), which offer functionalities such as internet connectivity and integration with other devices [Almehdhar et al., 2024]. The emergence of connected autonomous vehicles introduces complex optimization problems in cybersecurity defense, requiring advanced techniques from operations research (OR), artificial intelligence, and integrated systems to maintain both security integrity and operational efficiency.

The optimization of security frameworks for connected autonomous vehicles represents a critical application of OR in defense systems. As noted by Liu et al. [2017], these vehicles utilize sensor networks, data processing algorithms, and artificial intelligence to perform complex tasks with minimal human intervention, creating a multidimensional decision space where security and performance must be simultaneously optimized.

However, as these vehicles become more connected, new demands arise to ensure security and reliability [Almehdhar et al., 2024]. Since this connectivity also brings vulnerabilities that expose vehicles to cyberattacks, it can compromise everything from data privacy to physical vehicle control. In particular, the CAN (Controller Area Network) bus, widely used for internal communication between vehicle components, is a recurring target due to the absence of authentication and encryption mechanisms.

In this context, Intrusion Detection Systems (IDSs) play an essential role by analyzing data traffic patterns to identify deviations or anomalies that may indicate intrusion attempts [Korium et al., 2024]. These systems represent a practical application of OR principles, formulating the security problem as a multi-objective optimization where detection accuracy must be maximized while minimizing false alarms under computational resource constraints.

With the evolution of vehicular networks and technologies like 5G and IoT, IDS must be adaptable. Since IDS relies on machine learning and pattern analysis algorithms, the performance of these systems is directly related to the diversity, representativeness, and relevance of the data feeding the model. To accurately detect threats, it is essential that the dataset contains diverse attacks and behaviors [Sanaboina et al., 2023], ensuring the effectiveness and generalization of IDS in real-world scenarios.

Thus, this work aims to explore the Car-Hacking Dataset [Lee et al., 2017] through an operations research perspective. Our contribution differentiates itself by analyzing the intrusion detection problem as a multi-objective optimization challenge, where the goal is to maximize threat identification effectiveness while minimizing false alarm rates, considering the computational constraints of vehicular environments. Rather than proposing new datasets or focusing primarily on technical limitations, we provide a systematic quantitative analysis of intrusion detection as an optimization problem in defense systems. This operations research approach contributes to the development of more efficient security infrastructures for autonomous transportation networks, based on mathematical models for decision support in cyber risk scenarios.

2. Autonomous Vehicles and Cybersecurity

The development of autonomous vehicles marks a transformation in the transportation sector, enabling new forms of mobility that promise to enhance user safety, efficiency, and convenience. Equipped with advanced sensors and artificial intelligence systems, these vehicles can perform complex tasks such as navigation, environmental monitoring, and real-time decision-making, reducing the need for human intervention.

The classification of autonomous vehicles is based on levels of automation, ranging from 0 to 5, according to the definition by the Society of Automotive Engineers (SAE)¹. Figure 1 illustrates

¹ https://www.sae.org/standards/content/j3016_202104/

this classification.

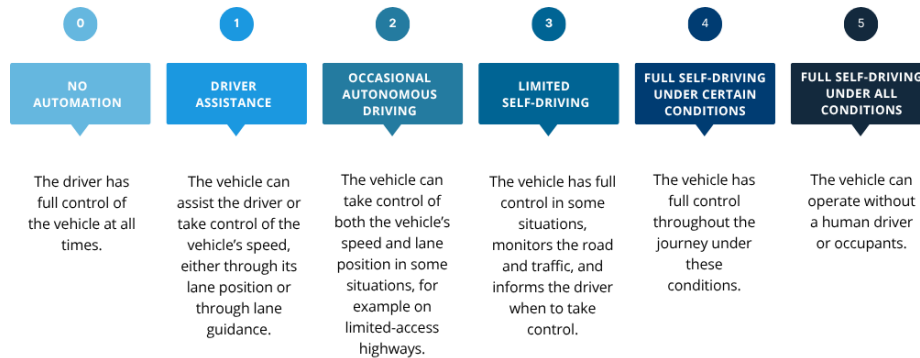


Figure 1: Representation of vehicle autonomy levels according to SAE (Society of Automotive Engineers) classification. Levels range from 0 (no automation) to 5 (full autonomy), indicating the degree of control and intervention in the vehicle. Adapted from [Kim, 2025].

At level 0, there is no automation, meaning the driver has full control, although the vehicle may offer basic assistance systems such as collision warnings. Level 5 represents full automation, where the vehicle can operate under all conditions and scenarios without any human intervention, enabling connectivity between the vehicle and other network elements [Silva et al., 2024]. However, this connectivity with devices, external networks, and infrastructures also makes them vulnerable.

Operations research provides essential mathematical frameworks for modeling the trade-offs between security, performance, and resource utilization in autonomous vehicular systems. These models help engineers and security specialists understand the complex decision boundaries between normal operations and potential security breaches, allowing for more efficient resource allocation in defensive measures.

Thus, cybersecurity in autonomous vehicles is a critical aspect, as failures in control systems can lead to severe consequences [Korium et al., 2024]. The internal systems of autonomous vehicles include multiple connection points, including electronic control units (ECUs) to communication interfaces like Wi-Fi, Bluetooth, and long-range networks susceptible to cyberattacks [Mansourian et al., 2023].

To protect the integrity of autonomous vehicles, security mechanisms are required to monitor and detect suspicious activities or anomalies in their internal and external networks. In this scenario, Intrusion Detection Systems (IDSs) emerge as a solution in threat identification.

3. Operations Research for Autonomous Vehicle Cybersecurity

Autonomous Driving Systems (ADSs) face significant cybersecurity challenges due to their reliance on interconnected networks and machine learning algorithms. Modern vehicles depend on electronic control units (ECUs) connected through networks like the Controller Area Network (CAN), which were designed primarily for efficiency rather than security [Koscher et al., 2010]. This architectural vulnerability creates a complex decision space where OR methodologies can provide systematic approaches to security optimization.

Security vulnerabilities in ADSs can be categorized into two main domains: infrastructure-based attacks targeting ECUs and algorithmic attacks exploiting machine learning components. ECU-based attacks take advantage of communication protocol weaknesses, potentially compromising critical vehicle functions [Korium et al., 2024]. Common attack vectors include infiltration through infotainment systems and manipulation of diagnostic interfaces like the On-Board Diagnos-

tics II (OBD-II) [Kang et al., 2018]. Countermeasures in the literature include distributed firewall systems and code authentication techniques that optimize the security-performance trade-off.

The machine learning components of ADSs introduce additional security concerns through adversarial machine learning. The perception and decision-making algorithms in ADSs can be compromised through poisoning attacks during training or evasion attacks during operation [Biggio and Roli, 2018]. For example, Boloor et al. [2019] demonstrated that physical modifications to road surfaces could manipulate ADS algorithms into unsafe steering behaviors, highlighting the need for robust detection systems.

Adversarial risk analysis provides an operations research framework for enhancing ADS security by modeling the interactions between defenders and attackers while accounting for uncertainty [Rios Insua et al., 2020]. This approach enables more effective decision rules for intrusion detection systems operating under threat scenarios [Naveiro et al., 2019], important when protecting critical vehicle functions from malicious interference that could affect passenger safety.

The effectiveness of intrusion detection systems depends significantly on the representativeness of the datasets used for training and testing [Sanaboina et al., 2023]. Operations research contributes valuable optimization techniques for dataset selection, feature extraction, and algorithm training that maximize detection accuracy while minimizing false positives. For datasets like the Car-Hacking Dataset examined in this work, OR methods help determine optimal classification thresholds that balance sensitivity to different attack types while maintaining computational efficiency in resource-constrained vehicular environments.

As autonomous vehicle technology advances, the quality and diversity of security datasets will play an increasingly critical role in developing robust protection systems. Operations research approaches that optimize the evaluation and utilization of these datasets represent a significant contribution to ensuring that intrusion detection systems can effectively identify and respond to emerging cybersecurity threats in the automotive domain.

4. Intrusion Detection Systems

Intrusion Detection Systems (IDSs) protect vehicular networks from attacks. These systems are designed to monitor data traffic, identify patterns that exhibit anomalies, and respond to potential threats before they cause significant damage [Abdulganiyu et al., 2023].

There are two main types of IDSs: signature-based and anomaly-based. Figures 2a and 2b illustrate the functioning of each type of IDS. Signature-based IDSs work by storing a set of patterns or signatures of known attacks in a database [Otoum and Nayak, 2021]. When new data packets arrive on the network, their characteristics are compared with the stored signatures. This approach is highly efficient in detecting previously documented attacks, as identification occurs accurately.

Moreover, the low false-positive rate makes this method reliable in environments where most threats have already been cataloged. However, it is not effective for detecting unknown attacks that lack pre-registered signatures. To maintain its effectiveness, this IDS relies on frequent updates to the database, which can be challenging.

On the other hand, anomaly-based IDSs continuously monitor network traffic and use techniques such as machine learning and statistical analysis to identify deviations from expected behavior [Otoum and Nayak, 2021]. These systems analyze the attributes of data packets, distinguishing normal traffic from malicious traffic based on learned patterns, being effective in detecting new attacks that do not have known signatures.

However, anomaly-based IDSs face challenges related to the accuracy of their algorithms. An ineffective model may generate false positives, flagging legitimate activities as threats, or false negatives, allowing attacks to go unnoticed [Abdulganiyu et al., 2023]. This can overwhelm administrators with unnecessary alerts or compromise network security.

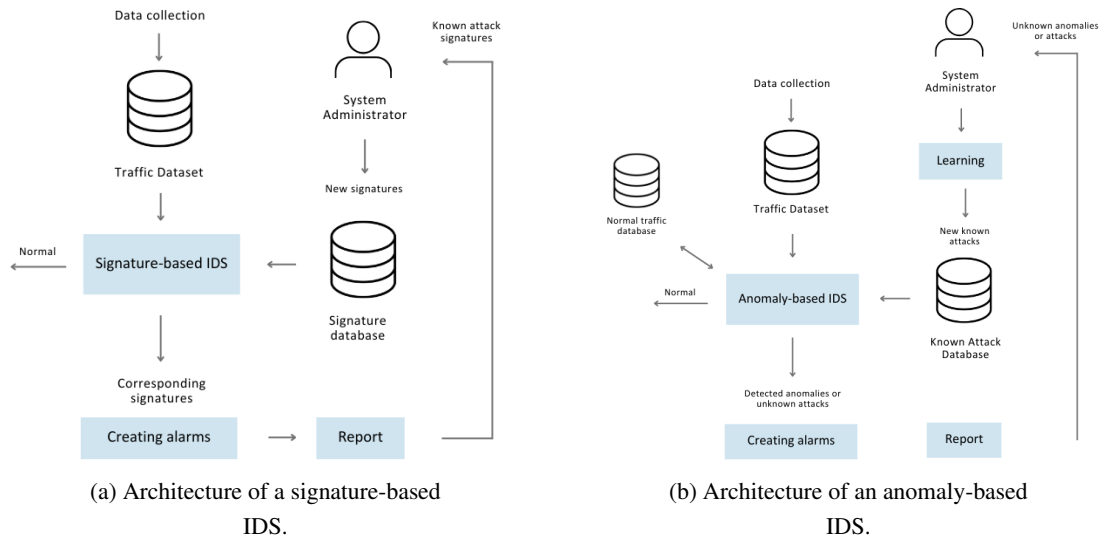


Figure 2: Comparison between the architectures of signature-based and anomaly-based IDS. Adapted from [Bangui et al., 2022].

The performance of this type of IDS depends directly on the quality of the algorithms used, the representativeness of the data used for training, and the system's ability to adapt to changes in network behavior over time [Sanaboina et al., 2023]. Additionally, the training process may require significant computational effort.

Thus, the effectiveness of Intrusion Detection Systems (IDSs) is directly related to their ability to identify and mitigate the various types of attacks that can compromise vehicle security [Abdulganiyu et al., 2023]. These attacks exploit vulnerabilities both in internal networks, such as the CAN bus, and in external communications, such as vehicular ad hoc networks (VANETs). Understanding the characteristics of these attacks is essential for designing more effective IDSs.

5. Controller Area Network (CAN) Bus

The Controller Area Network (CAN), developed by Bosch in the 1980s [Buscemi et al., 2023], is a widely used communication protocol for internal vehicle networks, enabling efficient and reliable data exchange between components like ECUs, sensors, and actuators. While it supports standardized messaging and adapts to modern Vehicle-to-Vehicle and Vehicle-to-Infrastructure connectivity demands [Mansourian et al., 2023], this new connectivity also brings security challenges.

The CAN bus operates based on the transmission of data packets organized into standardized frames, known as CAN data frames. As detailed by Buscemi et al. [2023], each frame includes fields such as SOF (start of transmission), arbitration (priority based on 11-bit or 29-bit ID), control (packet integrity), and data (0 to 8 bytes). CRC detects errors, ACK confirms receipt, and EOF marks the end of the message. Figure 3 shows the CAN frame structure, designed to prevent collisions and ensure efficient communication.

Description	SOF	Arbitration Field	Control Field	Data Field	CRC	ACK	EoF	Inter Frame Space
Length	1 bit	12 bits	6 bits	0 to 8 bytes	16 bits	2 bits	7 bits	3 bits

Figure 3: The structure of the CAN data frame. Adapted from [Verma et al., 2024].

The CAN protocol covers the first two layers of the OSI (Open Systems Interconnection)

reference model: the physical layer and the data link layer. At the physical layer, the CAN bus uses two twisted wires, called CAN High (CANH) and CAN Low (CANL). There are two main implementations of the CAN 2.0 physical layer: high-speed CAN (ISO 11898-2) [Standard, 2003] and low-speed CAN (ISO 11898-3) [Standard, 2006].

At the data link layer (ISO 11898-1) [Boland et al., 2021], communication on the CAN bus occurs through frames. Each ECU periodically sends frames without including source or destination addresses. When multiple ECUs transmit simultaneously, a collision occurs, resolved through an arbitration process. Frames with lower IDs have higher priority, this ensures that critical messages are transmitted first.

Data collection from the CAN bus is generally performed using devices known as CAN loggers, which are physically connected to the bus. These devices can be attached to the OBD-II connector, present in most modern vehicles [Kang et al., 2018], allowing the capture of messages transmitted by ECUs.

According to Buscemi et al. [2023], CAN data is stored in DBC files, which contain descriptions of the position, meaning, and format of signals transmitted on the bus. These files allow raw data to be interpreted in human-readable formats, facilitating analysis and the development of CAN-based applications. However, some manufacturers have started implementing firewalls between the OBD-II connector and the CAN bus.

The CAN bus, despite its efficiency, has vulnerabilities that compromise vehicle security, such as the lack of source and destination identification in packets [Buscemi et al., 2023], making it susceptible to spoofing attacks, injection of false messages and denial of service (DoS). Physical access through the OBD-II port allows manipulation of critical components, while wireless connections such as Bluetooth and WiFi can also be exploited to send malicious packets. Attackers can remotely disable the brakes or alter steering behavior, exposing weaknesses in modern systems [Miller and Valasek, 2014].

6. Related Work

The operations research approach to cybersecurity in autonomous vehicles has evolved significantly in recent years, with special focus on multi-objective optimization algorithms for intrusion detection and decision-making models under uncertainty. Han et al. [2022] applied multi-objective optimization techniques to balance detection accuracy and resource utilization in CAN network-based intrusion detection systems, formulating the problem as a mathematical programming model that considers computational constraints of vehicular ECUs.

Lee et al. [2023] provided an analysis of datasets for IDSs in the context of the CAN bus, categorizing attacks into spoofing, suspension, and masquerade. This work highlights the importance of proper mathematical formulation of the decision space for intrusion detection systems, demonstrating how different data representations affect the decision boundary of classifiers and the balance between false positives and true positives, a classic optimization problem in OR.

Awotunde et al. [2021] proposed an operations research approach for intrusion detection systems in industrial networks, using a hybrid model that combines rule-based feature selection and deep neural networks. Their mathematical formulation of the feature selection process as an optimization problem is particularly relevant to our work, as it demonstrates how operations research techniques can significantly reduce the dimensionality of the decision space without compromising model effectiveness.

To address the limitations of existing datasets, the ROAD dataset Verma et al. [2024] was introduced, offering over 3 hours of real vehicular traffic data with various attack types. From an OR perspective, this dataset enables the construction of more robust decision models through

the integration of metadata about vehicular activities and physical effects of attacks, allowing the application of optimization techniques that consider the uncertainty in real vehicular environments.

Lee et al. [2017] complemented this discussion by proposing OTIDS, an intrusion detection system based on the analysis of remote frames on the CAN bus. Using metrics such as response time and offset ratio as decision variables in their mathematical model, OTIDS demonstrates how proper formulation of the classification problem can significantly improve computational efficiency, enabling real-time detection of anomalies caused by attacks such as DoS, fuzzing, and masquerade.

Although the dataset associated with OTIDS includes real data, it has limitations such as inconsistencies in documentation and the absence of complete suspension of legitimate messages during masquerade attacks. Nevertheless, the method contributes by exploring temporal characteristics of CAN frames, applying sequential analysis techniques that can be modeled as Markovian decision processes, established approach in OR for sequential decision problems.

Our work differentiates itself by exploring the Car-Hacking Dataset [Lee et al., 2017] through an optimization and decision support perspective. Rather than focusing on limitations or proposing new datasets, we analyze how different classification algorithms, interpreted as optimization problem solvers, handle the balance between maximizing threat detection and minimizing false alerts, a classic trade-off problem in OR.

7. Methodology

This section describes the operations research approach employed to formulate and solve the intrusion detection problem using the Car-Hacking Dataset, presenting the decision optimization framework from problem structuring to solution evaluation.

The Car-Hacking Dataset [Lee et al., 2017] was selected as the decision space for our security optimization model. It contains CAN (Controller Area Network) messages captured during simulated attacks, comprising 17,558,462 messages distributed across five classification states: Normal (988,987 messages), DoS (3,665,771 messages, including 587,521 injected), Fuzzy (3,838,860 messages, with 491,847 injected), Gear Spoofing (4,443,142 messages, including 597,252 injected), and RPM Spoofing (4,621,702 messages, with 654,897 injected).

The decision variables in our security optimization model include 12 attributes from the dataset: Timestamp, CAN ID, DLC, byte values (DATA[0] to DATA[7]), and the binary decision outcome Flag, indicating whether a message is normal (R) or malicious (T). The operational data sources were: gear_dataset.csv, DoS_dataset.csv, Fuzzy_dataset.csv, RPM_dataset.csv, and normal_run_data.txt.

The decision space preprocessing involved linear transformations of the CAN_ID from hexadecimal to integers, binary encoding of the Flag column (R=0 for normal, T=1 for injected), and categorical expansion through an Attack_Type column to optimize the classification space. The preprocessing phase also implemented data integrity constraints by removing inconsistent and redundant entries, followed by dataset integration to enable a comprehensive optimization model.

The computational optimization experiments were executed using Python version 3.12, employing specialized libraries for mathematical modeling and optimization: NumPy (1.25.0) for numerical operations, Pandas (2.1.0) for data transformation, Scikit-learn (1.3.0) for implementing machine learning-based optimization models, Imbalanced-learn (0.11.0) for stochastic resampling using the oversampling method, Synthetic Minority Oversampling Technique (SMOTE), and TensorFlow (2.13.0) for neural network optimization.

The experimental design employed stratified resource allocation, dividing the decision space into training (80%) and testing (20%) partitions. A resource-constrained optimization approach was implemented through the Synthetic Minority Over-sampling Technique (SMOTE)[Chawla

et al., 2002], which algorithmically generates synthetic samples to balance the optimization objective across all attack classes. The validation set remained unmodified to ensure realistic evaluation of the model's operational effectiveness.

Four supervised learning algorithms were used: Decision Tree, Random Forest, Gradient Boosting, and CNN. The Decision Tree was configured with gini splitting criterion, maximum depth of 10, minimum samples to split a node (5), and minimum samples per leaf (2). For Random Forest used 100 trees, maximum depth of 15, and bootstrap enabled. In turn, the Gradient Boosting employed a learning rate of 0.1, 200 boosting stages, maximum depth of 8, and subsampling of 0.8. The CNN architecture consisted of three dense layers (128, 64, and 32 units) with relu activation, batch normalization, dropout (0.3), and a softmax output layer. Training was performed with the Adam optimizer (learning rate=0.001) and sparse_categorical_crossentropy loss function.

The results obtained were analyzed to verify whether the Car-Hacking Dataset is representative and suitable for use in intrusion detection systems, evaluating its ability to reflect real cyber attack scenarios. In addition, to evaluate the performance of each algorithm in classifying the different types of attacks.

8. Analysis Results of the Car-Hacking Dataset

The Car-Hacking Dataset is a dataset created to facilitate research and development of security methods aimed at automotive networks, especially those based on the Controller Area Network (CAN) protocol. This dataset was designed to simulate real-world scenarios of cyberattacks on vehicles, allowing researchers and developers to analyze anomalous behavior patterns and create effective solutions to detect and mitigate these threats.

The dataset is divided into normal and injected messages, representing the attacks, as illustrated in Figures 4 and 5.

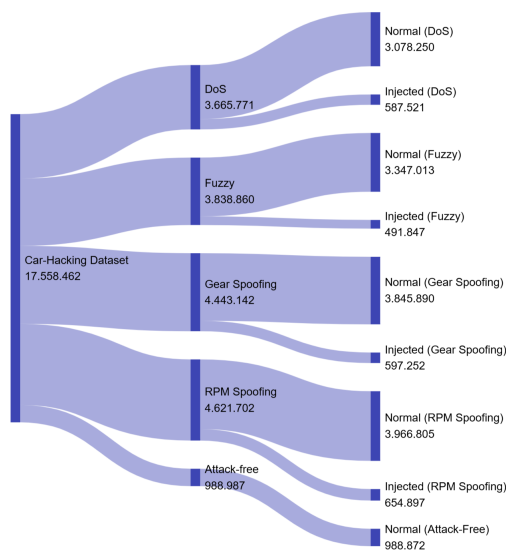


Figure 4: Distribution of messages in the Car-Hacking Dataset. The flows represent the types of attacks and normal traffic, divided between normal and injected messages.

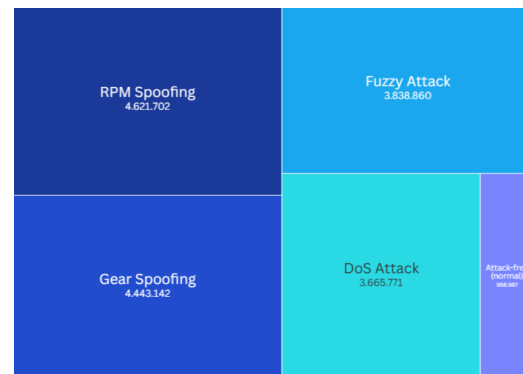


Figure 5: Distribution by types of attacks (DoS, Fuzzy, Gear Spoofing, RPM Spoofing) and normal traffic ('Attack-free') present in the dataset.

In the case of the denial-of-service attack (DoS attack), messages with the CAN identifier '0000' were injected every 0.3 milliseconds, with this identifier being the most dominant in the recorded traffic. For the fuzzy attack, messages with completely random CAN identifiers and data

values were injected every 0.5 milliseconds. In the spoofing attacks related to RPM and gear (spoofing attack), messages with specific CAN identifiers associated with this information were injected every 1 millisecond.

The dataset includes attributes describing CAN network messages captured during experiments. Key attributes are: Timestamp (time of each message in seconds), CAN ID (hexadecimal identifier), DLC (number of data bytes, 0-8), DATA[0] to DATA[7] (transmitted byte values), and Flag (R for normal messages, T for malicious ones).

Table 1: Overview of attack types and their characteristics in the dataset. The table provides details such as the timestamp interval, CAN ID format, output class (T for transmission and R for reception), and the total number of samples for each type.

Attack Type	Timestamp	CAN ID	Output Class	Samples
Normal	–	Random value	T	988.987
DoS	0.3 ms	Dominant value “0000”	R	587.521
Gear Spoofing	1 ms	Random hex value	R	597.252
RPM Spoofing	1 ms	Random hex value	R	654.897
Fuzzy	05 ms	Random hex value	R	491.847

8.1. Optimization Algorithms for Security Decision Making

Evaluating classification algorithms on the Car-Hacking Dataset provides a vision into their ability to detect and classify CAN messages. Performance was assessed using a multi-criteria decision framework incorporating precision, recall, and F1 score, as presented in Table 2.

Accuracy measures the proportion of instances correctly classified out of all instances predicted to be a given class. High accuracy indicates that the model is effective at minimizing false positives. For example, Random Forest achieved high accuracy for Gear Spoofing, suggesting that it rarely misclassified normal traffic or other types of attacks. This is important in scenarios where false alarms can lead to unnecessary interventions.

Recall, on the other hand, evaluates the proportion of correctly classified instances among all real instances of a given class. It reflects the model’s ability to identify true positives. CNN demonstrated 97% recall for both Gear Spoofing and normal traffic. This indicates its ability to detect nearly all instances of these classes, which is critical in intrusion detection systems where missing an attack can have serious consequences.

Regarding the F1-score, this metric combines precision and recall into a single metric, useful when dealing with imbalanced datasets. Gradient Boosting performed well for DoS attacks, achieving an F1-score of 80%. This suggests that it balances the identification of true positives and avoidance of false positives, making it suitable for detecting frequent attacks such as DoS.

Among the algorithms evaluated, CNN emerged as the best overall performer, achieving consistently high recall and F1-scores across multiple classes. While Gradient Boosting also showed strong results, particularly for DoS attacks, its performance was less consistent across other classes. Simpler models such as Decision Tree and Random Forest had specific strengths, such as RPM Spoofing and Gear Spoofing, respectively.

Therefore, these results highlight the importance of selecting metrics that are aligned with the objectives of the task. In the context of intrusion detection systems, recall is often prioritized to ensure that malicious activity is not missed, even if this comes at the cost of some false positives. Precision becomes critical in reducing false alarms, which can overwhelm security systems.

The algorithms evaluate the representativeness of the Car-Hacking Dataset for IDS, verifying its ability to simulate real-world cyberattacks. The results confirm its quality and potential to develop systems to detect and mitigate attacks.

Table 2: Classification report for different algorithms applied to the dataset after oversampling in the training set.

Algorithm	Class	Precision	Recall	F1-Score	Support
Decision Tree	DoS	0.78	0.72	0.75	235.331
	RPM Spoofing	0.74	0.85	0.79	188.847
	Gear Spoofing	0.65	0.55	0.60	164.991
	Fuzzy	0.70	0.62	0.66	129.290
	Normal	0.80	0.76	0.78	32.458
Random Forest	DoS	0.82	0.78	0.80	235.331
	RPM Spoofing	0.78	0.90	0.84	188.847
	Gear Spoofing	0.72	0.65	0.68	164.991
	Fuzzy	0.75	0.68	0.71	129.290
	Normal	0.85	0.82	0.83	32.458
Gradient Boosting	DoS	0.85	0.80	0.82	235.331
	RPM Spoofing	0.82	0.92	0.87	188.847
	Gear Spoofing	0.75	0.68	0.71	164.991
	Fuzzy	0.78	0.72	0.75	129.290
	Normal	0.88	0.85	0.86	32.458
CNN	DoS	0.90	0.82	0.85	235.331
	RPM Spoofing	0.88	0.94	0.91	188.847
	Gear Spoofing	0.80	0.72	0.76	164.991
	Fuzzy	0.82	0.75	0.78	129.290
	Normal	0.92	0.88	0.90	32.458

9. Conclusion

With the advancement of autonomous and connected vehicles, security has become a critical challenge that benefits from operations research (OR) principles to ensure system reliability and integrity. In this context, Intrusion Detection Systems (IDSs) represent an application of analytical modeling to security problems, using decision-making techniques to balance detection accuracy against false alarms while improving resource allocation for threat response.

The Car-Hacking Dataset serves as a structured decision space for developing OR-based security solutions, providing representative data that simulate real-world cyberattacks. Our application of multiple classification algorithms demonstrates how different decision models perform when evaluated through the quantitative metrics framework typical in OR. The CNN approach achieved superior performance across multiple attack vectors, highlighting how OR methodologies can effectively distinguish between normal and malicious patterns in vehicular network communication.

This work demonstrates the value of applying OR methodologies to cybersecurity challenges in transportation systems. By treating intrusion detection as a multi-criteria decision problem that balances detection sensitivity with specificity, we contribute to the advancement of more effective decision support systems for security operations. The systematic approach presented here shows how OR techniques can transform security data into actionable insights, managing the inherent trade-offs between comprehensive threat detection and efficient resource utilization in protecting critical transportation infrastructure.

Furthermore, it highlights the importance of high-quality datasets for the evolution of IDS in both internal and external vehicular networks. The representativeness and diversity of the data are essential factors to ensure that systems are adaptable to emerging threats and can operate reliably in real-world scenarios.

Finally, for future research in this OR application domain, we recommend expanding the

analysis framework by incorporating additional datasets to validate model performance across diverse scenarios. Alternative sampling techniques should be explored to address class distribution imbalances, improving the effectiveness boundary for less frequent attack types. Comparative analysis of different algorithms would also contribute to the OR literature by identifying which decision models offer the most efficient solutions for specific security threat categories in vehicular networks.

References

- Abdulganiyu, O. H., Ait Tchakoucht, T., and Saheed, Y. K. (2023). A systematic literature review for network intrusion detection system (ids). *International journal of information security*, 22 (5):1125–1162.
- Almehdhar, M., Albaseer, A., Khan, M. A., Abdallah, M., Menouar, H., Al-Kuwari, S., and Al-Fuqaha, A. (2024). Deep learning in the fast lane: A survey on advanced intrusion detection systems for intelligent vehicle networks. *IEEE Open Journal of Vehicular Technology*.
- Awotunde, J. B., Chakraborty, C., and Adeniyi, A. E. (2021). Intrusion detection in industrial internet of things network-based on deep learning model with rule-based feature selection. *Wireless Communications and Mobile Computing*, 2021:1–17.
- Bangui, H., Ge, M., and Buhnova, B. (2022). A hybrid machine learning model for intrusion detection in vanet. *Computing*, 104(3):503–531.
- Biggio, B. and Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84:317–331.
- Boland, H. M., Burgett, M. I., Etienne, A. J., and Stwalley III, R. M. (2021). An overview of canbus development, utilization, and future potential in serial network messaging for off-road mobile equipment. *Technology in Agriculture*.
- Bolloor, A., He, X., Gill, C., Vorobeychik, Y., and Zhang, X. (2019). Simple physical adversarial examples against end-to-end autonomous driving models. In *2019 IEEE International Conference on Embedded Software and Systems*, p. 1–7. IEEE.
- Buscemi, A., Turcanu, I., Castignani, G., Panchenko, A., Engel, T., and Shin, K. G. (2023). A survey on controller area network reverse engineering. *IEEE Communications Surveys & Tutorials*, 25 (3):1445–1481.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357. ISSN 1076-9757.
- Han, L., Duan, J., Wang, X., and Zhang, Y. (2022). Optimization of multi-objective intrusion detection systems for efficient resource allocation in automotive networks. *IEEE Transactions on Vehicular Technology*, 71(9):9632–9645.
- Kang, T. U., Song, H. M., Jeong, S., and Kim, H. K. (2018). Automated reverse engineering and attack for can using obd-ii. In *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, p. 1–7. IEEE.
- Kim, S.-K. (2025). Event driven technology intelligence. *Journal of Economy and Technology*, 3: 143–150.

- Korium, M. S., Saber, M., Beattie, A., Narayanan, A., Sahoo, S., and Nardelli, P. H. (2024). Intrusion detection system for cyberattacks in the internet of vehicles environment. *Ad Hoc Networks*, 153:103330.
- Koscher, K., Czeskis, A., Roesner, F., Patel, S., Kohno, T., Checkoway, S., McCoy, D., Kantor, B., Anderson, D., Shacham, H., and Savage, S. (2010). Experimental security analysis of a modern automobile. In *2010 IEEE Symposium on Security and Privacy*, p. 447–462. IEEE.
- Lee, H., Jeong, S. H., and Kim, H. K. (2017). Otds: A novel intrusion detection system for in-vehicle network by using remote frame. In *2017 15th Annual Conference on Privacy, Security and Trust (PST)*, p. 57–5709. IEEE.
- Lee, S., Choi, W., Kim, I., Lee, G., and Lee, D. H. (2023). A comprehensive analysis of datasets for automotive intrusion detection systems. *Computers, Materials & Continua*, 76(3).
- Liu, S., Tang, J., Zhang, Z., and Gaudiot, J.-L. (2017). Computer architectures for autonomous driving. *Computer*, 50(8):18–25.
- Mansourian, P., Zhang, N., Jaekel, A., and Kneppers, M. (2023). Deep learning-based anomaly detection for connected autonomous vehicles using spatiotemporal information. *IEEE Transactions on Intelligent Transportation Systems*, 24(12):16006–16017.
- Miller, C. and Valasek, C. (2014). A survey of remote automotive attack surfaces. *black hat USA*, 2014(2014):1–20.
- Naveiro, R., Redondo, A., Insua, D. R., and Ruggeri, F. (2019). Adversarial classification: An adversarial risk analysis approach. *International Journal of Approximate Reasoning*, 113:133–148.
- Otoum, Y. and Nayak, A. (2021). As-ids: Anomaly and signature based ids for the internet of things. *Journal of Network and Systems Management*, 29(3):23.
- Rios Insua, D., Baylon, C., and Vila, J. (2020). Security risk models for cyber insurance. *AXA Research Fund & ICMAT*.
- Sanaboina, S. P., Naik, M. C., and Rajiv, K. (2023). Examining the impact of artificial intelligence methods on intrusion detection with the nsl-kdd dataset. In *2023 First International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV)*, p. 1–7. IEEE.
- Silva, E., Soares, F., Souza, W., and Freitas, H. (2024). A systematic mapping of autonomous vehicle prototypes: Trends and opportunities. *IEEE Transactions on Intelligent Vehicles*, p. 1–27.
- Standard, I. (2003). Road vehicles-controller area network (can)-part 2: Highspeed medium access unit,â. *ISO 11898-2: 2003*.
- Standard, I. (2006). 11898-3: Road vehicles-controller area network (can)-part 3: Lowspeed, fault-tolerant, medium-dependent interface. *International Organization for Standardization (www.iso.org)*.
- Verma, M. E., Bridges, R. A., Iannacone, M. D., Hollifield, S. C., Moriano, P., Hespeler, S. C., Kay, B., and Combs, F. L. (2024). A comprehensive guide to can ids data and introduction of the road dataset. *PLoS one*, 19(1):e0296879.