

A Multi-class Intrusion Detection System for Cyber Security Education in Automotive Industry

Mirko De Vincentiis¹, Anibrata Pal¹, Azzurra Ragone¹ and Michele Scalera¹

¹University of Bari Aldo Moro, Department of Computer Science, Via Edoardo Orabona 4, Bari, Italy

Abstract

Connected electronic components within vehicles can be exploited by cyber attackers if not properly protected. Controller Area Network (CAN), the standard protocol used by the in-vehicular components to communicate among themselves, lacks security features for data protection. We need methodologies and solutions to increase cybersecurity awareness in the automotive industry to identify and protect vehicles from attacks that can exploit these security lacks. To reach this goal, this paper proposed a methodology to increase cybersecurity education in which education starts from the university using innovative research methodologies and then proposing strategies that could help the automotive industry. The proposed strategy adopts a multi-class Intrusion Detection System to identify CAN attacks.

Keywords

Automotive, IDS, Cybersecurity, ADAS, Education

1. Introduction

The increasing number of Electronic Control Units (ECUs), such as Advanced Driver Assistance Systems (ADAS), Bluetooth, and Infotainment Systems, make modern vehicles technological but potentially vulnerable to cyber-attacks. In particular, these components exchange information using different standard protocols, where Controller Area Network (CAN) is the most used due to its high resilience to electromagnetic interference and low cost [1, 2]. Since the CAN protocol does not implement encryption techniques, researchers have found several attacks that can be conducted by exploiting these vulnerabilities. For example, The security research lab of Tencent Keen Security Lab found several vulnerabilities in a Mercedes model. In particular, they captured the Service Set Identifier (SSID) and the passphrase transmitted by the head unit to the T-Box component using the CAN bus as plaintext [3].

To increase the security of the CAN protocol and the vehicle itself, several works have adopted the use of Machine Learning (ML) using a binary classification, but that does not provide the typology of attack. The use of ML algorithms that adopt multi-class strategies can be useful in supporting the automotive industries to understand not only the attack used by an adversary but also to create strategies to help the drivers [4].


For these reasons, the paper presents an Intrusion Detection System (IDS) aiming to detect cyber attacks in real-time and to improve cyber security education in the automotive industry.

IS-EUD 2023: 9th International Symposium on End-User Development, 6-8 June 2023, Cagliari, Italy

✉ mirko.devincentiis@uniba.it (M. De Vincentiis); anibrata.pal@uniba.it (A. Pal); azzurra.ragone@uniba.it (A. Ragone); michele.scalera@uniba.it (M. Scalera)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

The idea is to support the automotive industry in being able to reconstruct the attack kill chain and understand the impact of the attack on other components [5, 6]. Considering the vehicle as a collection of software, hardware, electronic, and mechanical components, it is necessary to understand how an attack could affect each of them and especially to identify security flaws from the different elements [7]. In addition, the vehicles can also communicate with external components to establish a connection between each other (Vehicle-to-Vehicle) or communicate with components located aside the roads (Vehicle-to-Infrastructure).

The paper is organized as follows: Section 2 describes the related works; Section 3 explains the proposed methodology about cyber security education in the automotive field; Section 4 discussed the CAN protocol and the dataset used to validate the model briefly before showing the experiments and results; and finally Section 5 presents the conclusion and future works.

2. Related Work

The next-generation vehicles will be subject to new types of attacks that could compromise not only the safety of the driver but also the reputation of the automotive company. Many of these attacks can be conducted because the in-vehicle protocols do not implement cybersecurity mechanisms [8]. In particular, since the CAN is the most used in-vehicle protocol and does not implement cryptographic and authentication mechanisms, researchers demonstrated that it is vulnerable to Denial-of-Service (DoS) and injection attacks (such as fuzzy and spoofing) [9, 10, 11, 12]. In August 2021, the ISO/SAE 21434 "Road vehicles - Cyber Security engineering" [13] was released to augment the security of the vehicles by enforcing security standards and recommendations. The document proposed a generic framework regarding the requirements for Cyber Security risk management for next-generation vehicles. Since it does not provide a concrete design methodology, it is necessary to provide solutions.

On the other hand, most of the research works define ML or Deep Learning (DL) techniques as solutions to increase security in the CAN protocol. Since the ECUs have limited computational resources, traditional ML algorithms can be used instead of DL models [14]. The Random Forest model showed good results in identifying attacks on the CAN bus [14, 15, 16].

It is also important to increase cybersecurity awareness to communicate and counter cyber-attacks [17, 18]. With this consideration, this paper proposed a methodology considering a cybersecurity framework to increase cybersecurity awareness in the automotive industry. As a strategy to identify CAN attacks, a multi-class IDS was proposed using a Random Forest model.

3. Cyber Security Education in Automotive Industry

Cyber Security education plays an important role in helping data protection in different industries and public offices by training and preparing security specialists. Although industries conduct cyber security training from time to time, this is primarily provided by the Universities through research and collaborations. Figure 1 presents the overall goal and the strategies we followed regarding cyber security education.

The principal goal is to educate and train industry resources to raise automotive risk awareness. To achieve this, we divide the goal into objectives and subsequently into actionable

Goal	Objective	Strategy
1. Raise awareness about risks in automotive	1.1 Improve knowledge of risks and vulnerabilities in automotive	1.1.1 Promote cyber security awareness through collaboration between academia and industry
		1.1.2. Promote strategies for training the company to manage cyber risk
	1.2 Promote the use of cyber security resources and tools	1.2.1 Integration of different stakeholders to understand the risks of different components involved in automotive industry
		1.2.2 Improve cyber security-focused activities

Figure 1: Goal and Strategy of Cyber Security Education in Automotive Industry

strategies. To implement the strategies we follow the National Initiative for Cybersecurity Education (NICE) framework [19]. The principal building blocks of the NICE framework are the Tasks, Knowledge, and Skills (TKS) statements, which incorporate agility, flexibility, interoperability, and modularity as important attributes. In our case, the NICE framework (Figure 2) can be utilized to impart **Knowledge**, education and training on automotive cyber security domain, and **Skill**, technical and business skills to remediate automotive vulnerabilities, to perform the **Task** of implementing effective and accurate security measures to remediate automotive cyber-attacks.

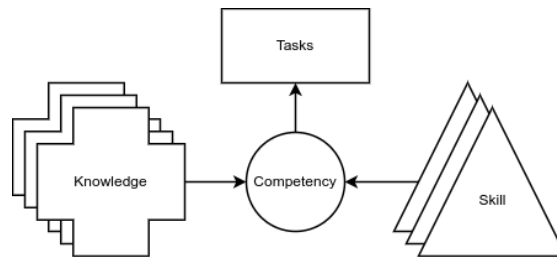


Figure 2: Adaptation of NICE framework

With reference to the proposed educational strategies and the NICE framework, Figure 3 shows the proposed methodology, where the essential elements for cyber security education start from the University and terminate in the automotive industry with a possible stakeholder involved to manage the automotive cyber-attacks. Firstly, the university investigates the possibilities of new automotive cyber attacks on next-generation vehicles. For example, information about new attack typologies that may be used to attack vehicles can be used as necessary domain knowledge (Figure 1, **Strategy - 1.1.1**) and can be used in education and training. In particular, understanding what kind of cyber attack is occurring in the vehicle is essential to increase the competence and the skills of the stakeholders that operate in the automotive field. The use of Machine Learning algorithms could be a solution for automotive companies to manage cyber risk and attacks (Figure 1, **Strategy - 1.1.2**).

A multi-class Intrusion Detection System (IDS) has been proposed as a solution to identify the

attacks and protect the in-vehicle network. The proposed solution could be extremely useful in identifying the typology of the attacks and helping the automotive industry to adapt response solutions based on the attack type. Considering Figure 3, for example, the Security Operation Center (SOC) Analyst (a professional responsible for a company’s cybersecurity and security operations), based on the attack that occurred in the vehicle, can make decisions on how to respond to the attack based on his knowledge and skill [20]. Imparting the knowledge and understanding of these outcomes across the organization could ensure cyber security awareness among resources like developers, testers, and architects (Figure 1, **Strategy - 1.1.2 and 1.2.1**). Furthermore, the industry can use the knowledge and skills necessary for cyber security to develop new competencies in the employees to spread and equip specialists for specific cyber security tasks (Figure 1, **Strategy - 1.2.2**). The automotive industry, apart from the University led research and development programs, also should strictly adhere to the ISO/SAE 21434 ”Road vehicles - Cyber Security engineering” [13] regarding the secure development process and hardware component. Thus, the industry can assimilate innovative cyber security solutions to learn to protect particular components from cyber attacks by continuous monitoring, surveys, and deployment of novel onboard security systems.

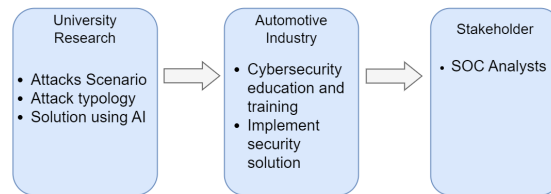


Figure 3: Proposed methodology for Cyber Security education.

4. Experiment

This section presents a brief summary of the Controller Area Network, attack typology, and the dataset for testing the multi-class IDS. Consequently, it presents the multi-class Intrusion Detection System as a solution for cybersecurity education in the automotive industry, and the classification results obtained thereof.

4.1. Controller Area Network

The CAN protocol [21] allows the ECUs to exchange messages between them. To send the information about a message, the CAN protocol uses the Data frame, which is subdivided into different frames that are *Identifier (ID)* it is used for the arbitration phase; *Data Length Code (DLC)* specify the length of the payload sent by an ECU; *Data* it contains the information about the message. The arbitration phase in the CAN protocol is used to avoid collision when two or more ECUs send messages simultaneously.

4.2. Attack Typology

There are three identified attack scenarios that can be conducted in internal networks of vehicles: **Denial-of-Service (DoS)** comprises sending high-priority messages to block communication with other nodes; **Fuzzy**, sends a spoofed random CAN ID, causing a change in vehicle behavior; and **Spoofing** consists of sending messages of a specific ID [22, 12].

4.3. Car-Hacking Dataset

The Hacking and Countermeasure Research Lab (HCRL) published a dataset called Car-Hacking Dataset¹ [23, 22] that contains real CAN messages from a Hyundai's YF Sonata logged using an OBD-II port. The authors of the dataset also performed three attacks: Denial of Service (DoS), Fuzzy, and Spoofing. The Car-Hacking Dataset is subdivided into four datasets: DoS, Fuzzy, Spoofing Revolutions Per Minute (RPM), and Spoofing Gear. The authors inject the CAN ID '0316' for the Spoofing RPM. Instead, the CAN ID is '043f' for the Spoofing Gear. Each of these contains normal and attack messages.

The data attributes present in the datasets are **Timestamp** is the recorded time; **CAN ID**, the identifier of the message in hexadecimal form; **DLC** indicates the number of bytes from 0 to 8; **DATA** the payload in hexadecimal form from DATA[0]-DATA[7]; **Flag T** represents an injected message while **R** a normal message.

4.4. Multi-class IDS solution

The Multi-class IDS is tested on CAN data, a popular in-vehicle network protocol. A *Random Forest* model was trained using the state-of-the-art *Car-Hacking Dataset* dataset containing real CAN messages from real vehicles. The experiments were conducted on a device with an Intel Core i7-11800H processor and 32 GB of RAM using Python 3 and the Scikit-learn library [24].

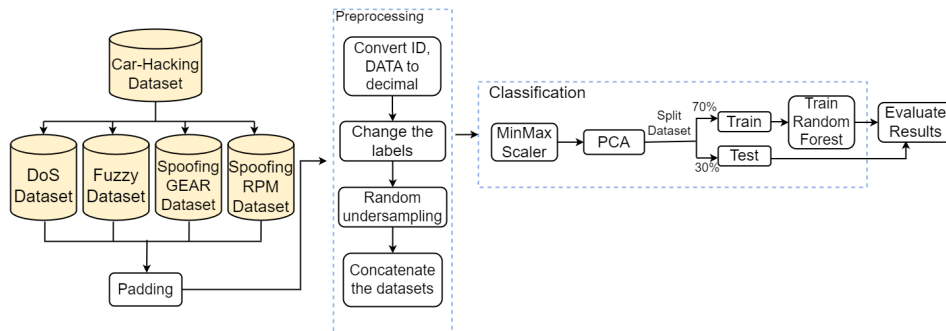


Figure 4: Proposed methodology

Figure 4 shows the proposed IDS methodology. The DoS, Fuzzy, Spoofing RPM, and Spoofing GEAR datasets were padded in the pre-processing phase. The padding process adds a '00' value where the DATA attribute is Not a Number (NaN). This process avoids removing examples from the datasets. After this phase, the DATA and CAN ID were transformed from hexadecimal to

¹<https://ocslab.hksecurity.net/Datasets/car-hacking-dataset>

Table 1

Results obtained using the Random Forest classifier for the concatenated dataset in multi-class approach.

Model	Attack	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	Normal		99.99	99.99	99.99
	DoS		100	100	100
	Fuzzy	99.99	99.98	99.99	99.98
	Spoofing GEAR		100	100	100
	Spoofing RPM		100	100	100

decimal. Then, the datasets were relabeled by mapping T with 1 (for DoS), 2 (for Fuzzy), 3 (for Spoofing GEAR), 4 (for Spoofing RPM), and R with 0 for no attack message. Since the datasets are unbalanced, the random under-sampler was used for the majority class, which is the normal class. Finally, the processed dataset was concatenated to make a multi-class classification. After the pre-processing phase, the attributes were scaled using the MinMaxScaler² in a range of [0, 1], and the Principal Component Analysis (PCA) was used. The default parameters were used for Random Forest. Finally, the concatenated dataset was split into 70% for training and 30% for testing. Accuracy, Precision, Recall, and F1-Score were used to evaluate the Random Forest model, and the results are shown in Table 1. The proposed approach obtained 100% for the DoS, Spoofing GEAR, and Spoofing RPM. For the Fuzzy attack, instead, the model reached 99.98% due to the randomness of the CAN ID and DATA, which leads the model to classify the normal messages as Fuzzy sometimes.

5. Conclusion and Future Work

Cybersecurity is critical in the automotive field because if an attack occurs in a vehicle, it could compromise the driver's life. For this reason, education about the cybersecurity of the automotive industry is important to protect cars from cyber attacks. To reach this goal, this paper proposed a methodology considering the NICE framework where the education starts from the university using innovative research methodologies and then proposing strategies that could help the automotive industry improve vehicle security. This strategy uses a multi-class IDS with a Random Forest model to identify three important CAN attacks: DoS, Fuzzy, and Spoofing. The results show that this model reached good results with a state-of-the-art dataset and could improve the security of the in-vehicle network. In future work, we plan to create a Random Forest model that can be deployed on an ECU to detect CAN attacks, and to analyse the dynamical behaviour of data [25, 26].

Acknowledgments

This study has been partially supported by the following projects: SSA (Secure Safe Apulia – Regional Security Center, Codice Progetto 6ESURE5) and KEIRETSU (Codice Progetto V9UFIL5)

²<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>

funded by "Regolamento regionale della Puglia per gli aiuti in esenzione n. 17 del 30/09/2014 (BURP n. 139 suppl. del 06/10/2014) TITOLO II CAPO 1 DEL REGOLAMENTO GENERALE "Avviso per la presentazione dei progetti promossi da Grandi Imprese ai sensi dell'articolo 17 del Regolamento"; and SERICS (Security and Rights In the CyberSpace - PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

References

- [1] A. Venturi, D. Stabili, F. Pollicino, E. Bianchi, M. Marchetti, Comparison of machine learning-based anomaly detectors for controller area network, in: 2022 IEEE 21st International Symposium on Network Computing and Applications (NCA), volume 21, 2022, pp. 81–88. doi:10.1109/NCA57778.2022.10013527.
- [2] V. S. Barletta, D. Caivano, A. Nannavecchia, M. Scalera, Intrusion detection for in-vehicle communication networks: An unsupervised kohonen som approach, *Future Internet* 12 (2020). URL: <https://www.mdpi.com/1999-5903/12/7/119>. doi:10.3390/fi12070119.
- [3] T. S. K. Lab, Experimental Security Assessment of Mercedes-Benz Cars, <https://keenlab.tencent.com/en/2021/05/12/Tencent-Security-Keen-Lab-Experimental-Security-Assessment-on-Mercedes-Benz-Cars/>, 2021.
- [4] V. S. Barletta, D. Caivano, C. Catalano, M. De Vincentiis, A. Pal, Machine learning for automotive security in technology transfer, in: *Information Systems and Technologies - WorldCIST 2023 Volume 1*, 2023.
- [5] F. Tommasi, C. Catalano, U. Corvaglia, I. Taurino, Mineralert: an hybrid approach for web mining detection, *Journal of Computer Virology and Hacking Techniques* 18 (2022) 333–346.
- [6] C. Catalano, A. Chezzi, M. Angelelli, F. Tommasi, Deceiving ai-based malware detection through polymorphic attacks, *Computers in Industry* 143 (2022) 103751.
- [7] V. S. Barletta, D. Caivano, A. Nannavecchia, M. Scalera, A kohonen som architecture for intrusion detection on in-vehicle communication networks, *Applied Sciences* 10 (2020). URL: <https://www.mdpi.com/2076-3417/10/15/5062>. doi:10.3390/app10155062.
- [8] A. Martínez-Cruz, K. A. Ramírez-Gutiérrez, C. Feregrino-Uribe, A. Morales-Reyes, Security on in-vehicle communication protocols: Issues, challenges, and future research directions, *Computer Communications* 180 (2021) 1–20. URL: <https://www.sciencedirect.com/science/article/pii/S0140366421003297>. doi:<https://doi.org/10.1016/j.comcom.2021.08.027>.
- [9] M. Bozdal, M. Samie, I. Jennions, A survey on can bus protocol: Attacks, challenges, and potential solutions, in: 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), 2018, pp. 201–205. doi:10.1109/iCCECOME.2018.8658720.
- [10] B. Groza, P.-S. Murvay, Security solutions for the controller area network: Bringing authentication to in-vehicle networks, *IEEE Vehicular Technology Magazine* 13 (2018) 40–47. doi:10.1109/MVT.2017.2736344.
- [11] E. Aliwa, O. Rana, C. Perera, P. Burnap, Cyberattacks and countermeasures for in-vehicle networks, *ACM Computing Surveys* 54 (2021) 1–37. doi:10.1145/3431233.

- [12] D. Stabili, L. Ferretti, M. Andreolini, M. Marchetti, Daga: Detecting attacks to in-vehicle networks via n-gram analysis, *IEEE Transactions on Vehicular Technology* 71 (2022) 11540–11554.
- [13] I. I. S. Organisation, ISO/SAE DIS 21434 Road vehicles—cybersecurity engineering, 2021.
- [14] R. Gundu, M. Maleki, Securing can bus in connected and autonomous vehicles using supervised machine learning approaches, in: 2022 IEEE International Conference on Electro Information Technology (eIT), IEEE, 2022, pp. 042–046.
- [15] L. Yang, A. Moubayed, I. Hamieh, A. Shami, Tree-based intelligent intrusion detection system in internet of vehicles, in: 2019 IEEE global communications conference (GLOBECOM), IEEE, 2019, pp. 1–6.
- [16] A. Alfardus, D. B. Rawat, Intrusion detection system for can bus in-vehicle network based on machine learning algorithms, in: 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE, 2021, pp. 0944–0949.
- [17] M. Khader, M. Karam, H. Fares, Cybersecurity awareness framework for academia, *Information* 12 (2021). URL: <https://www.mdpi.com/2078-2489/12/10/417>. doi:10.3390/info12100417.
- [18] V. S. Barletta, D. Caivano, M. De Vincentiis, A. Magrì, A. Piccinno, Quantum optimization for iot security detection, in: V. Julián, J. Carneiro, R. S. Alonso, P. Chamoso, P. Novais (Eds.), *Ambient Intelligence—Software and Applications—13th International Symposium on Ambient Intelligence*, Springer International Publishing, Cham, 2023, pp. 187–196.
- [19] R. Petersen, D. Santos, M. C. Smith, K. A. Wetzel, G. Witte, Workforce framework for cybersecurity (NICE framework), NIST Special Publication 800-801 (2020). URL: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-181r1.pdf>. doi:10.6028/NIST.SP.800-181r1.
- [20] V. S. Barletta, D. Caivano, M. D. Vincentiis, A. Ragone, M. Scalera, M. Á. S. Martín, V-SOC4AS: A Vehicle-SOC for Improving Automotive Security, *Algorithms* 16 (2023). URL: <https://www.mdpi.com/1999-4893/16/2/112>. doi:10.3390/a16020112.
- [21] Bosch, CAN Specification Version 2.0, Robert Bosch GmbH, Postfach 50 (1991).
- [22] H. M. Song, J. Woo, H. K. Kim, In-vehicle network intrusion detection using deep convolutional neural network, *Vehicular Communications* 21 (2020) 100198.
- [23] E. Seo, H. M. Song, H. K. Kim, Gids: Gan based intrusion detection system for in-vehicle network, in: 2018 16th Annual Conference on Privacy, Security and Trust (PST), 2018, pp. 1–6. doi:10.1109/PST.2018.8514157.
- [24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [25] M. Angelelli, E. Ciavolino, P. Pasca, Streaming generalized cross entropy, *Soft Computing* 24 (2020) 13837–13851.
- [26] M. Angelelli, Tropical limit and a micro-macro correspondence in statistical physics, *Journal of Physics A: Mathematical and Theoretical* 50 (2017) 415202.