

Intra-Vehicular Network Security Datasets Evaluation

Achref Haddaji¹ ^a, Samiha Ayed² and Lamia Chaari Fourati³

¹*National School of Electronics and Telecommunications of Sfax, Tunisia*

²*LIST3N-ERA, University of Technology of Troyes, France*

³*Digital Research Center of Sfax (CRNS), SM@RTS (Laboratory of Signals, systems, aRtificial Intelligence and neTworkS), Sfax University, Tunisia*

Keywords: Vehicular Networks, Intra-Vehicular Networks, Security, Datasets, Cyber-Attacks, Artificial Intelligence.


Abstract: Vehicular networks are more and more connected to the outside world. Therefore they became highly vulnerable to different cyber-attacks by being an easy target. Consequently, intra-vehicular networks' cybersecurity risk is raised too. As a solution, Artificial Intelligence (AI) based solutions were proposed to overcome these issues. On the other hand, their effectiveness relies mainly on the existing sources and datasets to ensure the networks' security. However, there is a significant challenge to overcome: the studies of the existing datasets of intra-vehicular network security. To tackle this issue, this paper examines and assesses existing intra-vehicular network security datasets. In addition, we comprehensively provide a detailed resource on the existing datasets and elaborate a comparative study. This paper also presents outstanding research discussions on dataset preprocessing, usability, and strength points to guide and help researchers.

1 INTRODUCTION

1.1 General Context

In the last decade, the rapid adoption of intelligent vehicles (Shokravi et al., 2020), also known as connected vehicles, has revolutionized their networks and security. These advanced vehicles employ sophisticated technologies based on Artificial Intelligence (AI) (Haddaji et al., 2022) that cooperate with intelligent vehicle components (e.g., sensors). AI interferes with different tasks, such as communicating with other vehicles, infrastructure, and the internet, enhancing user safety, comfort, and performance efficiency. However, connected vehicles' advancement in the automotive environment has been returned explicitly to In-vehicle networks (Rajapaksha et al., 2023). Intelligent cars rely heavily on in-vehicle networks where many functions linked to sensors and processors within the vehicle are used. They enable various electronic systems and control components to communicate and exchange data. These data include features like adaptive cruise control, lane departure warning, and blind spot detection. As in-vehicle networks become more intricate and integrated, cyber-

attackers have a lot of entry points. Vulnerabilities can be exploited via the vehicle's systems and the variety of interactions between them, such as the CAN bus (Jichici et al., 2022), the primary communication channel by the majority of in-vehicle systems. An adversary who obtains access to the CAN bus may be able to manipulate the data sent between the various vehicle systems, causing the vehicle to behave erratically or even become uncontrollable. The vehicle's wireless interfaces, such as Bluetooth, Wi-Fi, or cellular networks, are potential attack vectors. An attacker with access to these interfaces may be able to implement remote attacks, such as injecting malicious code or commands into the vehicle's systems. In addition, physical access to the car, such as through the diagnostic port or other external interfaces, can facilitate attacks. Denial-of-service (DoS) attacks (Shah et al., 2022), remote code execution, and physical attacks (Duo et al., 2022) that enable an attacker to control the vehicle's steering, stopping, or acceleration are examples of attacks demonstrated on in-vehicle networks. Overcoming these issues, protecting the safety and privacy of intelligent vehicles and their occupants, and preventing cyber-attacks require ensuring the security of in-vehicle networks.

^a  <https://orcid.org/0000-0002-0388-9840>

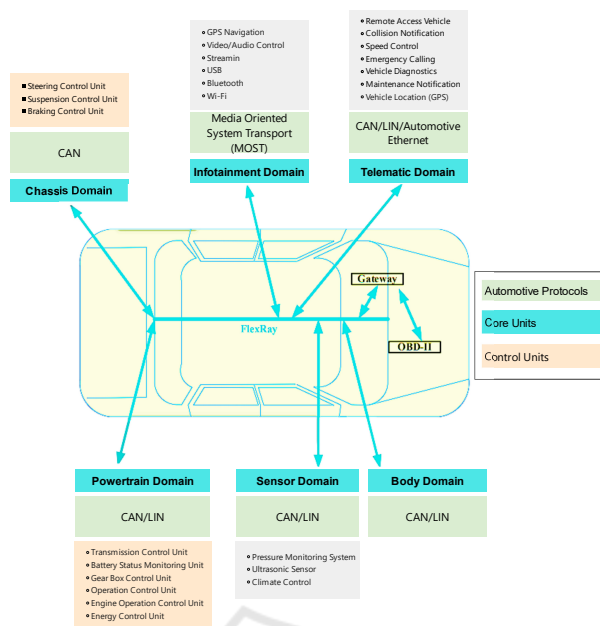


Figure 1: Intra-Vehicle Network Architecture: Automotive Protocols and Units.

1.2 Motivation and Problematic

Since in-vehicle networks are vulnerable to both internal and external attacks, it is crucial to have robust security measures in place to protect these systems from cyber risks. To address these issues, experts in the area of vehicular security have created many techniques and strategies for in-vehicle network security based on AI. AI-based solutions establish different Machine Learning (ML) and Deep Learning (DL) algorithms to track and analyze network traffic, identify abnormalities and suspicious behavior, and instantly respond to threats. Specifically, researchers showed considerable interest in intra-vehicular networks attacks detection. Moreover, recent advancements in AI could assist the vehicle by identifying and repairing the systems and network flaws before attackers can take advantage of them. However, AI solutions development and innovation are related to the available resources (e.g., simulations, datasets, and experiment information). Meanwhile, there is a considerable need to have more resources to validate these approaches. Therefore, the need for available datasets and open resources represents a big challenge for vehicular network security (intra-vehicular networks specifically). Therefore, this challenge created a need for research studies to assess and survey public datasets for intra-vehicular network security. Indeed, a limited number of studies concentrate on vehicular network security datasets (intra-vehicular datasets). This fact might significantly affect and decrease the efficacy of security solutions. To tackle

this challenge, the primary objective of this paper is to address, review and analyze the currently available datasets in vehicular network security. In addition, the value of this work is represented by providing a comprehensive exposition of the different existing datasets utilized in AI-based solutions to enhance vehicular communication security.

1.3 Contributions and Outline

This paper includes a more in-depth exploration of intra-vehicular network security datasets. Therefore, the major contributions are as follows:

- Present an overview of intra-vehicular networks principles, protocols, and security issues.
- Assess and evaluate the available intra-vehicular networks security datasets.
- Highlight the preprocessing phase and its major steps and characteristics.
- Discuss the available datasets' norms or usage, benefits, and limitations.

The remainder of this paper is organized as follows: First, section 2 presents an overview of intra-vehicular networks. Then, Section 3 list and review the existing intra-vehicular network security datasets. Section 4 provides a discussion and identifies the potential of each dataset, followed by a conclusion in Section 5.

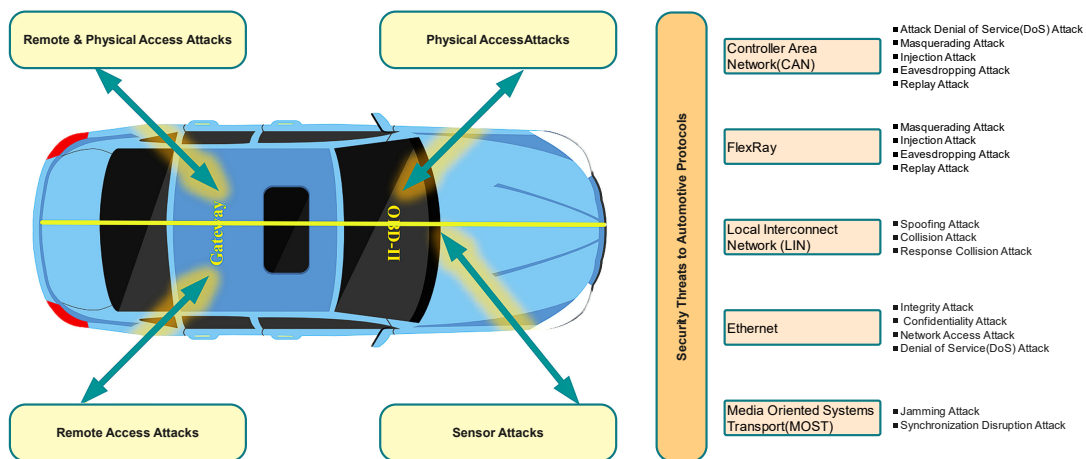


Figure 2: Classification of Possible Security Threats: Entry Surfaces and Automotive Protocols.

2 INTRA-VEHICULAR NETWORKS: OVERVIEW

This section presents a brief background knowledge about Intra-Vehicular Networks and their various related security concerns.

2.1 Intra-Vehicular Networks Preliminaries

It consists of several core components (e.g., gateways, sensors, actuators, etc.) distributed in various units, namely the sensor domain, chassis domain, telematics domain, powertrain domain, etc. the communication between these components is based on the usage of protocols that play a major role. As described in (Rathore et al., 2022), there are three major classification types of intra-vehicular architectures. The first type is based on the central gateway known as distributed electrical and electronics (E/E). Meanwhile, the second architecture is multiple operational domains linked through a central gateway, known as the domain-centralized electricals and electronics (E/E). The last classification type is known as future E/E architecture or zonal architecture. It consists of a centralized high-performance computing unit (HPCU) that aims to reduce the complexity of previously existing two architectures. Figure 1 illustrated the general architecture of intra-vehicular networks.

2.2 Intra-Vehicular Networks Automotive Protocols

Connected vehicles are equipped with an advanced sensor platform capable of transmitting a high num-

ber of signals internally. This sensor data is processed by approximately 70 Electronic Control Units (ECUs) interconnected within the vehicle. The intra-vehicular network enables the exchange of data between sensors, ECUs, and actuators, which is crucial for the vehicle's proper functioning. Therefore, the primary communication systems involve substantial use of five intra-vehicular networks protocols (Aksu and Aydin, 2022): (1) Local Interconnection Network (LIN), (2) Controller Area Network (CAN), (3) FlexRay, (4) Ethernet, and (5) Media Oriented Systems Transport (MOST). Each protocol has advantages and disadvantages (See Table 1). For example, LIN offers a low communication speed and is suited for applications that do not demand precise time performance, such as battery monitoring and window actuator control. In addition, LIN has a limited fault-tolerance capability. On the other hand, FlexRay, Ethernet, and MOST offer greater bandwidth than LIN, making them ideal for time-sensitive and bandwidth-intensive applications. FlexRay, for instance, is employed in safety systems such as steering angle sensors and safety radar. In contrast, Ethernet and most are commonly utilized in the infotainment system and ECU flash interface. Due to its low cost, mature tool networks, and acceptable noise-resistance and defect tolerance performance, CAN is the most popular network due to its low cost, proper noise-resistance performance, and fault tolerance (Al-Jarrah et al., 2019).

2.3 Intra-Vehicular Networks Security Issues

Intra-vehicular networks have been known as an easy target for attackers owing to their complexity and

Table 1: Classification of Intra-vehicular Networks Communication Protocols.

Network Speed	Bandwidth	Topology	Max Supported Nodes	Advantages	Limitations
CAN	25 Kbps – 1 Mbps	Star, Ring, Linear bus	30	High reliability, low cost	Limited bandwidth, vulnerable to attacks
LIN	25 Kbps – 1 Mbps	Liner bus	16	Bus Low cost, low power	Limited data rate and distance
FlexRay	Up to 10 Mbps	Star, Linear bus, hybrid	22	High reliability, high bandwidth	Higher cost, limited interoperability
Ethernet	Up to 100 Mbps	Star, Linear bus	Depends on Switch ports	High bandwidth, scalable	Higher cost, high power consumption
MOST	Up to 150 Mbps	Ring	64	High bandwidth, low latency	Limited distance, higher cost

open issues caused by the existing vulnerabilities. These security issues are described as follows:

- Lack of adequate bus protection, leaving messages vulnerable to interception, modification, and fabrication, and lacking necessary protections such as confidentiality, integrity, authenticity, and non-repudiation.
- Authentication issues, allowing unauthorized re-programming of ECU firmware, posing safety risks and enabling control over critical components.
- Protocol implementation issues, where deviations from safety rules and guidelines compromise system reliability and safety.
- Data leakage issues, enabling unauthorized access to private vehicle data, violating privacy and compromising security.
- Misuse of protocols, leveraging mechanisms like bus arbitration and fault detection to launch disruptive attacks on the network.

However, intra-vehicular network attacks might sneak from different entry points (e.g., Sensors, Physical surfaces, and remote mediums). Moreover, automotive protocols also present an easy target too. Figure 2 depicts some attacks and a graphical representation of the entry points.

3 INTRA-VEHICULAR NETWORKS DATASETS

As vehicles become more connected and rely on ECUs, the importance of intra-vehicular network security is becoming increasingly obvious. Indeed, intra-vehicular networks rely on the protocols. Within this context, this section examines the available intra-vehicular network datasets. In addition, this section presents the most important phase, which is the pre-processing phase. Moreover, Table 2 evaluates all the

above datasets and analyzes their advantages and limitations.

3.1 Existing Datasets

3.1.1 Car-Hacking Dataset

The car hacking dataset consists of CAN packets collected from the OBD-II terminal. Each CAN packet is defined by three important features: CAN ID, which represents the CAN packet's identifier, DATA[0] to DATA[7], which defines the packet's 8 bytes; and the flag, which admits two possible values, T and R. (T: inject packet and R: normal packet). Normal traffic and three forms of attack are included in this dataset. (1) DoS attack: CAN ID = 0X000 DoS packets are injected every 0.3 milliseconds. (2) Flexible attack: Every 0.5 milliseconds, random ID and DATA values are injected. (3) Spoofing Attack (RPM/gear): It injects RPM and gear-related CAN ID packets every millisecond.

3.1.2 OTIDS Dataset

OTDIS(Lee et al., 2017) represents the Offset Ratio and Time Interval based Intrusion Detection System which is a novel IDS based on the timing of remote frame responses. The basic strategy consisted of transmitting remote frame requests for a given ID, measuring how long it took an ECU to respond, and determining whether this delay was unusual; the idea was that a compromised ECU under the control of an adversary would respond with an unusual delay. This dataset is generated by collecting CAN packets via the OBD-II port. It comprises both normal transmissions and DoS attacks with a CAN ID of "0X000." It also includes fuzzy and impersonation attacks. The CSV files associated with fuzzy and impersonation attacks do not indicate whether a packet is normal.

Table 2: Intra-vehicular Communication Datasets: Comparison.

Dataset	Ref/Year	Objective	Attacks	Nature of Data	Format	Label	Protocol
Car-Hacking	(Seo et al., 2018)	Intrusion Detection	DoS, Fuzzy, Spoofing	Real	CSV	Yes	CAN protocol
OTIDS	(Lee et al., 2017)	Intrusion Detection	DoS, Fuzzy, Impersonation	Real	CSV	No	CAN protocol
Survival	(Han et al., 2018)	Intrusion Detection	Flooding, Fuzzy, Malfunction	Real	CSV	Yes	CAN protocol
SynCAN	(Hanselmann et al., 2020)	Intrusion Detection	Suspension, Fabrication, Masquerade	synthetic	CSV	No	CAN protocol
TU/e v2	(Dupont et al., 2019)	Intrusion Detection	DoS, Fuzzy, Diagnostic, Replay, Suspension	synthetic	CSV	No	CAN protocol
ROAD	(Verma et al., 2020)	Intrusion Detection	Masquerade, Fabrication targeted ID, Accelerator	Real	CSV	Yes	CAN protocol
CrySyS	(Chiscop et al., 2021)	Intrusion Detection	Plateau attack, Continuous change attack, Playback attack, Suppression attack, Flooding attack	Real data and synthetic attacks	CSV	No	CAN protocol, GPS
SIMPLE	(Foruhandeh et al., 2019)	Intrusion Detection	Dominant Impersonation, Complete Impersonation	Real	NA	Yes	CAN protocol
Bi	(Bi et al., 2022)	Intrusion Detection	Dos, Fuzzy, Ulterior Fuzzy, Replay	Real	NA	No	CAN protocol

3.1.3 Survival Dataset

HCRL released two datasets derived from three distinct vehicles: the "Kia Soul," "Hyundai Sonata," and "Chevrolet Spark." One of the datasets contains normal driving records, while the other contains driving records that are anomalous due to three attack scenarios: flooding, fuzzy, and malfunction. These attacks consisted of implanting attack messages every 20 seconds for five seconds, capturing each threat for 25-100 seconds. The dataset was used to develop a survival analysis-based detection model (Han et al., 2018) capable of identifying anomalies in in-vehicle networks. Survival analysis is a statistical technique that focuses on the timing of an event.

3.1.4 SynCAN

The SynCAN Dataset (Hanselmann et al., 2020) is a standard for comparing and contrasting various CAN Intrusion Detection Systems (IDS) using multiple attack scenarios in the signal space. It consists of a training dataset and six testing datasets, each of which contains columns for labels, IDs, time, and signal values. The following files contain the six datasets used for testing: *testnormal.zip* contains only normal data with a label of 0 for evaluating IDS performance on unmodified data. Other files include *testplateau.zip*, in which a signal's value remains constant over time, and *testcontinuous.csv*, in which a signal's value progressively deviates from its actual value. The dataset also includes *testplayback.zip*, in which a signal is overwritten with a recorded time series of the same signal, *testsuppress.zip*, in which messages of a specific ID are absent from the CAN traffic due to an attacker preventing an ECU from sending messages, and *testflooding.zip*, in which an attacker sends messages of a specific existing ID at a high frequency to the CAN bus. This dataset is intended to facilitate the

unsupervised training and evaluation of IDS on both normal and aberrant data.

3.1.5 TU/e v2 Dataset

In their study, the authors in (Dupont et al., 2019) suggested a framework for evaluating intrusion detection systems (NIDSs) for Controller Area Network (CAN) networks. They gathered data from two vehicles, Opel Astra and Renault Clio, and a CAN bus prototype that they constructed to generate their dataset. Additionally, they utilized Kia Soul data from the car-hacking dataset. The dataset is available online at the Eindhoven University of Technology Lab (TUE Security Group (Group, 2019)). The authors introduced a sequence of attacks against the prototype to generate attack datasets and then simulated these attacks on vehicles. They randomly injected ten packets with CAN IDs greater than 0x700 to perform a diagnostic attack. Next, they carried out two fuzzing attacks, which included injecting ten packets with unknown CAN IDs and altering the payload of ten frames with a valid CAN ID. They also performed a replay attack by injecting an arbitrary packet that occurred 30 times in the dataset and modifying the timestamp to send the packets ten times quicker than usual. To simulate a DoS attack, messages with a CAN ID of 0x000 were sent at a rate of four packets per millisecond to replace all messages within a 10-second period. Finally, the authors simulated a suspension attack by deleting all messages containing a specific CAN ID over a 10-second period.

3.1.6 ROAD Dataset

The ROAD dataset (Verma et al., 2020) comprises 12 ambient captures with approximately 3 hours of ambient data and 33 attack captures with a total runtime of around 30 minutes. All the data was collected from

a single vehicle whose make and model are not disclosed. All the data was collected from a single vehicle whose make and model are not disclosed. Three categories are used to classify the attacks that were recorded on the dataset. The first category is the fusing attack in which the authors injected frames with random IDs every 0.005s. The second category consists of targeted ID fabrication & masquerade attacks. The authors used the flam delivery technique for targeted ID fabrication, in which a message is injected immediately after a legitimate message containing the target ID is seen. For the masquerading attack version, the authors deleted the legitimate target ID frames preceding each injected frame to simulate a masquerade attack. Finally, accelerator attacks are an additional category in which the attack uses a vulnerability particular to the vehicle make/model, compromising the ECUs.

3.1.7 CrySyS Dataset

The CrySyS Lab created the publically available dataset (Chiscop et al., 2021) for the SECREDas project. It includes seven captures and one extended driving scenario trace, along with 20 message IDs and varying signal numbers. In addition to the dataset, the authors created a signal extractor and attack generator script that can modify CAN messages in various ways, including changing to constant or random values, modifying with delta or increment/decrement values, or switching to increment/decrement values. In addition, the attack generator can be used to simulate attacks by substituting a selected signal in the CrySyS traces with a constant value.

3.1.8 SIMPLE Dataset

The SIMPLE dataset (Foruhandeh et al., 2019) is a collection of public data obtained by capturing CAN messages from two vehicles, a 2016 Nissan Sentra and a 2011 Subaru Outback, through the OBD-II interface with a Tektronix DPO 3012 oscilloscope. During each round, the vehicles were driven for approximately 40 minutes, including local and highway traffic. The dataset includes more than 16,000 frames. Each frame in this data set comprises six parts: CAN high voltage samples, CAN low voltage samples, time interval, sample rate, decoded bits, and message ID. In this data set, Hill climbing-style attacks are included.

3.1.9 Bi's Dataset

The dataset proposed in (Bi et al., 2022) is generated from various driving situations. It is used with CAN

traffic acquired from the test vehicle's daily commute route. The vehicle's route included three different scenarios: country roads, highways, and congested city roads. The dataset had 29213281 messages and contained seven days of CAN traffic gathered during commuter driving. The dataset included challenging road conditions like slippery, congested, rainy, and foggy roads. The authors injected anomalous data into the CAN bus of the test vehicles using data injection equipment. They used four attack models in the vehicle's stationary state and driving state to generate the attack dataset. The attack messages included DoS attacks, fuzzy attacks, ulterior fuzzy attacks, and replay attacks.

3.2 Data Pre-Processing

The preprocessing phase is characterized by different steps inside which differ from one dataset to another. They highly depend on the dataset format, type, size, and features, among others. On the other hand, this phase shares many similarities applied to the dataset, being a general structure without specificities. Before discussing the characteristics of existing datasets, it is essential to understand in-vehicle data clearly. The CAN bus data is widely researched due to its primary usage as a data source. The CAN frame structure consists of seven fields: Start of the frame (SOF), Arbitration Field (identifier and RTR), Control Field, Data Field, CRC Field, Acknowledge Field, and End of Frame. These fields serve various purposes such as initiating transmission, prioritizing messages, verifying successful reception, transmitting data, ensuring message integrity, confirming successful receipt, and signaling frame termination.

However, Intra-vehicular network datasets are generated by simulating ECUs vehicles injecting CAN messages in a controlled environment. Therefore, data might be collected from a single vehicle or multiple vehicles. Hence, the data preprocessing phase comes directly after the data acquisition. This major phase comprises different steps, such as normalization, data cleaning, and feature encoding, which may be common for many datasets. There are other steps, such as feature selection, resizing, and format conversion, may need to be customized based on the unique characteristics of each dataset. Regarding feature encoding, it is important to convert qualitative values, such as "normal" or "attack," to integer values. For binary classification, the values should be altered to "0" and "1," while for multi-class classification, the values should range from "1" to "n," where "n" denotes the number of classes. Meanwhile, for CAN ID data, hexadecimal values should

Table 3: Inter-vehicular Communication Datasets: Recommendation.

Dataset	DATA source	Data type	Best Attack Detected	Worst Attack Detected	Recommendation
Car-Hacking	Multiple Vehicles	Standard CAN data	ID attacks	DoS attack	****
OTIDS	Single vehicle	Standard CAN data	Fuzzy attack	Masquerading attack	**
Survival	Multiple vehicles	Standard CAN data	Fuzzy attack	Malfunction attack	***
SynCAN	Single vehicle	Signal	Masquerading attack	-	****
TU/e v2	Multiple vehicles	Standard CAN data	Suspension, Masquerading	DoS	****
ROAD	Single vehicle	Standard CAN data and Signal data		masquerading	****
CrySyS	Single Vehicle	Standard CAN data, GPS data	Masquerading	-	**
SIMPLE	Single vehicle	Signal	Complete impersonation	Dominant impersonation	***
Bi	Single vehicle	Standard CAN data	Dos, Fuzzy,	Replay	**

be converted to decimal values using specific functions (such as the "hex2dec" function). For the data field, spaces between bytes should be removed using the "gsub" function, and the hexadecimal data value should then be converted to decimal integers (the "Rmpfr" function could be used as a function). Finally, Some datasets may have different file formats that need to be converted to a common format before preprocessing can be performed.

4 DISCUSSION

Available intra-vehicular network security datasets are very limited. Moreover, the existing datasets focus mainly only on CAN bus protocol and do not give attention to the other protocols. The car hacking dataset is the most used in the context of CAN IDS literature. The attack recordings in this dataset comprise a large number of instances per attack. The ID attacks present the gear and RPM functions in this dataset. However, the attack simulations are not occurred when the car is driven, which makes the test data different from the training data. In addition, data are in different formats, which is undesirable. On the other hand, these available datasets study redundant attacks (e.g., DoS attacks and fuzzy attacks). The OTIDS is the only dataset that provides a slightly stealthier version of spoofing IDS in normal traffic. Therefore, this dataset can be used for identity spoofing-based systems. In addition, it is the only dataset with remote frames and responses. However, this dataset is not recommended for many reasons. First, the injection intervals need to be clarified and explained in the documentation. Then, Although the attack was labeled as a masquerade attack in the paper, it may not meet the criteria of a true masquerade attack since the legitimate node's message transmission was not suspended. Finally, remote frame requests and responses result in minor timing variations, which may pose a challenge when testing and training a timing-based detector. The Survival dataset includes attacks on three vehicles that can have a real effect on the vehicle. Therefore, this dataset is a good choice for a simple timing-based detector. However,

similar to car hacking and OTIDS datasets, the attacks are basic and simple to detect. Furthermore, the amount of data offered for each vehicle in ambient captures is only 60-90 seconds, which is inadequate to ensure reliable training and to examine false positive rates. SynCAN is one of the most known datasets that is based on the signal. It is quite similar to ROAD dataset and SIMPLE dataset. This dataset compromises attacks that target a single signal and the full 64-bit data field, which allows for testing very advanced IDS-based signals. However, this dataset can not be used by the IDS that use the CAN data in the standard format (IDs with data fields).

Researchers that would simulate diagnostic protocol attacks could use TU/e v2 dataset. This dataset is the only dataset that includes suspension attacks in standard CAN data. However, this dataset is not used to simulate the DoS attacks. In addition, the data generation process needs to be clarified for this dataset. Finally, accessing information about the injected packets and their timing is complicated because the attack labels are stored in an unstructured text file.

The ROAD dataset is one of the recent datasets that treat the limitation of the previous datasets. It provides different types of fuzzy attacks. Furthermore, it is the only dataset in which the attacks are physically verified. In addition, this dataset provided both CAN data and CAN signal. However, the masquerading attacks rely on a small amount of simulation. In addition, this dataset could not provide a high resolution for testing time-based detectors because the time stamps are accurate only to 100us.

Crysys dataset is a good dataset to simulate and detect masquerading attacks. In addition, this dataset gives a clear idea about the injection time. Crysys is the only dataset that describes the driver's actions during data capturing. The only limitation of this dataset is that the attacks are added after the post-processing, which can affect vehicle functions. Finally, there are two datasets, namely SIMPLE and Bi's, which are private. The access is not available for public users, and they need the permission of the creator to use them. These aforementioned datasets are analyzed and compared based on different metrics such as their source, type, best and worst detected attacks using

this dataset, respectively, and usability recommendation (See Table 3).

5 CONCLUSION

Both sectors, including research and the industry, have shown incredible concerns about vehicular network security. Therefore, intra-vehicular network security needs to be addressed as well. In accordance with the current solutions, studying intra-vehicular security datasets will provide a strong base for the research and development to acquire valuable enhanced solutions. This paper is devoted to presenting a comprehensive study of various intra-vehicular network security datasets and their related quality measures. In addition, this study addresses the major phase of datasets, which is preprocessing. Moreover, it examines the available existing datasets and presents their impact through comparative analyses that show their benefits and limitations.

REFERENCES

- Aksu, D. and Aydin, M. A. (2022). Mga-ids: Optimal feature subset selection for anomaly detection framework on in-vehicle networks-can bus based on genetic algorithm and intrusion detection approach. *Computers & Security*, 118:102717.
- Al-Jarrah, O. Y., Maple, C., Dianati, M., Oxtoby, D., and Mouzakitis, A. (2019). Intrusion detection systems for intra-vehicle networks: A review. *IEEE Access*, 7:21266–21289.
- Bi, Z., Xu, G., Xu, G., Tian, M., Jiang, R., and Zhang, S. (2022). Intrusion detection method for in-vehicle can bus based on message and time transfer matrix. *Security and Communication Networks*, 2022.
- Chiscop, I., Gazdag, A., Bosman, J., and Biczók, G. (2021). Detecting message modification attacks on the can bus with temporal convolutional networks. *arXiv preprint arXiv:2106.08692*.
- Duo, W., Zhou, M., and Abusorrah, A. (2022). A survey of cyber attacks on cyber physical systems: Recent advances and challenges. *IEEE/CAA Journal of Automatica Sinica*, 9(5):784–800.
- Dupont, G., Den Hartog, J., Etalle, S., and Lekidis, A. (2019). Evaluation framework for network intrusion detection systems for in-vehicle can. In *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVEx)*, pages 1–6. IEEE.
- Foruhandeh, M., Man, Y., Gerdes, R., Li, M., and Chantem, T. (2019). Simple: Single-frame based physical layer identification for intrusion detection and prevention on in-vehicle networks. In *Proceedings of the 35th annual computer security applications conference*, pages 229–244.
- Group, T. S. (2019). Eindhoven university of technology.
- Haddaji, A., Ayed, S., and Fourati, L. C. (2022). Artificial intelligence techniques to mitigate cyber-attacks within vehicular networks: Survey. *Computers and Electrical Engineering*, 104:108460.
- Han, M. L., Kwak, B. I., and Kim, H. K. (2018). Anomaly intrusion detection method for vehicular networks based on survival analysis. *Vehicular Communications*, 14:52–63.
- Hanselmann, M., Strauss, T., Dormann, K., and Ulmer, H. (2020). Canet: An unsupervised intrusion detection system for high dimensional can bus data. *Ieee Access*, 8:58194–58205.
- Jichici, C., Groza, B., Ragobete, R., Murvay, P.-S., and Andreica, T. (2022). Effective intrusion detection and prevention for the commercial vehicle sae j1939 can bus. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):17425–17439.
- Lee, H., Jeong, S. H., and Kim, H. K. (2017). Otids: A novel intrusion detection system for in-vehicle network by using remote frame. In *2017 15th Annual Conference on Privacy, Security and Trust (PST)*, pages 57–5709. IEEE.
- Rajapaksha, S., Kalutarage, H., Al-Kadri, M. O., Petrovski, A., Madzudzo, G., and Cheah, M. (2023). Ai-based intrusion detection systems for in-vehicle networks: A survey. *ACM Computing Surveys*, 55(11):1–40.
- Rathore, R. S., Hewage, C., Kaiwartya, O., and Lloret, J. (2022). In-vehicle communication cyber security: challenges and solutions. *Sensors*, 22(17):6679.
- Seo, E., Song, H. M., and Kim, H. K. (2018). Gids: Gan based intrusion detection system for in-vehicle network. In *2018 16th Annual Conference on Privacy, Security and Trust (PST)*, pages 1–6. IEEE.
- Shah, Z., Ullah, I., Li, H., Levula, A., and Khurshid, K. (2022). Blockchain based solutions to mitigate distributed denial of service (ddos) attacks in the internet of things (iot): A survey. *Sensors*, 22(3):1094.
- Shokravi, H., Shokravi, H., Bakhary, N., Heidarrazaei, M., Rahimian Kolor, S. S., and Petru, M. (2020). A review on vehicle classification and potential use of smart vehicle-assisted techniques. *Sensors*, 20(11):3274.
- Verma, M. E., Iannacone, M. D., Bridges, R. A., Hollifield, S. C., Kay, B., and Combs, F. L. (2020). Road: The real ornl automotive dynamometer controller area network intrusion detection dataset (with a comprehensive can ids dataset survey & guide). *arXiv preprint arXiv:2012.14600*.