future internet

MDPI

Review

Toward Generative AI-Based Intrusion Detection Systems for the Internet of Vehicles (IoV)

Isra Mahmoudi [1], Djallel Eddine Boubiche [1,*], Samir Athmani [1], Homero Toral-Cruz [2,*] and Freddy I. Chan-Puc [2]

[1] LEREESI Laboratory, HNS-RE2SD, Batna 05000, Algeria; mahmoudi.isra@hns-re2sd.dz (I.M.); samir.athmani@hns-re2sd.dz (S.A.)
[2] Department of Sciences and Engineering, University of Quintana Roo, Chetumal 77019, Mexico; freddy@uqroo.edu.mx
* Correspondence: de.boubiche@hns-re2sd.dz (D.E.B.); htoral@uqroo.edu.mx (H.T.-C.)

Abstract

The increasing complexity and scale of Internet of Vehicles (IoV) networks pose significant security challenges, necessitating the development of advanced intrusion detection systems (IDS). Traditional IDS approaches, such as rule-based and signature-based methods, are often inadequate in detecting novel and sophisticated attacks due to their limited adaptability and dependency on predefined patterns. To overcome these limitations, machine learning (ML) and deep learning (DL)-based IDS have been introduced, offering better generalization and the ability to learn from data. However, these models can still struggle with zero-day attacks, require large volumes of labeled data, and may be vulnerable to adversarial examples. In response to these challenges, Generative AI-based IDS—leveraging models such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Transformers—have emerged as promising solutions that offer enhanced adaptability, synthetic data generation for training, and improved detection capabilities for evolving threats. This survey provides an overview of IoV architecture, vulnerabilities, and classical IDS techniques while focusing on the growing role of Generative AI in strengthening IoV security. It discusses the current landscape, highlights the key challenges, and outlines future research directions aimed at building more resilient and intelligent IDS for the IoV ecosystem.

Keywords: IoV; intrusion detection system; generative AI; machine learning; deep learning; transformers; GANs; VAEs

check for updates

Academic Editors: Olusola Tolulope Odeyomi and Temitayo Olowu

Received: 17 May 2025
Revised: 9 July 2025
Accepted: 15 July 2025
Published: 17 July 2025

Citation: Mahmoudi, I.; Boubiche, D.E.; Athmani, S.; Toral-Cruz, H.; Chan-Puc, F.I. Toward Generative AI-Based Intrusion Detection Systems for the Internet of Vehicles (IoV). *Future Internet* **2025**, *17*, 310. https://doi.org/10.3390/fi17070310

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Internet of Vehicles represents a network that links vehicles, users, and intelligent devices to the Internet. It expands the Internet of Things (IoT) into advanced transportation by implementing next-generation technologies and employing mobile vehicles as carriers for information perception. The IoV allows vehicles to access resources such as cloud storage and computing, aiming to increase transportation efficiency and optimize the driving experience [1].

As the Internet of Vehicles continues to expand, it is transforming transportation through seamless connectivity between vehicles, infrastructure, and cloud systems. However, this increased interconnectivity also introduces significant security challenges, making IoV networks a prime target for cyber threats. To mitigate these risks, Intrusion Detection Systems powered by ML and DL have been developed to detect and counter cyberattacks.

These AI-driven IDS analyze vast amounts of network data to identify malicious activities and safeguard vehicle communication.

Despite their advancements, ML- and DL-based IDS face key challenges, including high false positive rates, difficulty adapting to evolving attack strategies, and computational inefficiencies in real-time detection. Traditional models often struggle to capture complex attack patterns over time due to limitations in sequential data processing. Moreover, centralized IDS architectures pose scalability and security concerns, making them less effective in large-scale IoV environments.

To address these limitations, researchers are turning to Generative AI-based IDS, which offer enhanced capabilities in detecting and responding to cyber threats. Generative models, such as Transformers and Large Language Models (LLMs), excel in analyzing sequential data and capturing temporal correlations in network traffic. These models improve IoV security by generating realistic attack scenarios for robust training, reducing false alarms, and adapting to emerging threats with greater precision. As a result, Generative AI provides a promising alternative to conventional IDS, paving the way for more resilient and intelligent cybersecurity solutions in IoV networks.

1.1. Existing Reviews and Our Contributions

Intrusion Detection Systems have become a vital component in the security architecture of the IoV, given the increasing attack surface introduced by vehicle connectivity and autonomous functionalities. Several survey papers have examined IDS development in the context of IoV and internal vehicle networks, particularly highlighting the role of artificial intelligence (AI) in enhancing detection accuracy and adaptability.

Man et al. [1] provide a broad overview of AI-based IDS in IoV, covering both internal vehicle networks and external communication links such as V2V and V2I. Their taxonomy spans misuse-based, anomaly-based, and hybrid models, and emphasizes the importance of anomaly detection for unknown threat mitigation. However, the review remains general in scope; lacks technical depth regarding deep learning architectures; and does not explore Generative AI-based IDS techniques, such as GANs or transformer models, which are increasingly applied in cybersecurity for detecting novel, unseen attacks.

In the context of in-vehicle networks (IVNs), several surveys have explored AI-based intrusion detection approaches. Almehdhar et al. [2] provide a focused review of DL-based IDS models for CAN-based systems, proposing a taxonomy based on detection strategies, attack types, datasets, and AI techniques. Their work introduces key emerging methods such as federated learning and reinforcement learning, though it lacks coverage of transformer-based or Generative AI methods.

Complementing this, Rajapaksha et al. [3] present one of the most detailed and recent surveys dedicated exclusively to AI-based IDS for IVNs (2016–2022). They introduce a novel taxonomy that categorizes models based on the features used and the type of AI technique (ML, DL, hybrid), and include various attack types. Notably, this survey also reviews the security of AI models, potential threats against them, and practical deployment steps. While this work does mention GANs and even a few uses of Transformers for sequence modeling, the coverage is still sparse and exploratory rather than systematic. Generative AI is not treated as a distinct category in their taxonomy, and the latest LLM-based or fine-tuned transformer models are not considered. A comprehensive taxonomy in Table 1 is introduced below, organizing existing AI-based IDS surveys in vehicular networks.

Table 1. Comparative taxonomy of AI-based IDS surveys in vehicular networks.

Survey	Time Span	ML-Based IDS	DL-Based IDS	Generative AI-Based IDS			Network Focus
				GANs-IDS	Transformer-IDS	VAEs-IDS	
Man et al. [1]	---	✓	Limited	x	x	x	IVNs + V2X
Almehdhar et al. [2]	2017–2024	✓	✓	x	x	x	IVNs
Rajapaksha et al. [3]	2016–2022	✓	✓	x	x	x	IVNs
This Survey	2015–2024	✓	✓	✓	✓	✓	IVNs + V2X

Despite the growing interest in AI-based IDS for IVNs and broader IoV systems, a systematic investigation into the role of Generative AI remains missing from the literature. Existing reviews either focus narrowly on traditional ML/DL techniques or treat generative methods—such as GANs or Transformers—as isolated case studies without integrating them into a broader security framework. No prior work provides a dedicated taxonomy or structured analysis of Generative AI-based IDS across both internal (IVNs) and external (V2X) vehicular networks. Given the rise of data-driven attacks and the complexity of vehicular communication, this gap significantly limits the understanding of generative models' full potential in securing next-generation transportation systems.

To bridge this gap, this survey provides the first in-depth analysis of Generative AI-based Intrusion Detection Systems for the Internet of Vehicles. Unlike prior reviews that overlook or briefly mention generative models, this work focuses specifically on the application of GANs and Transformer-based architectures in vehicular security. It offers a comprehensive taxonomy of generative IDS approaches, categorizing them based on the type of generative model, deployment context (in-vehicle vs. external networks), and targeted attack types. In addition, the survey critically analyzes how these models handle core challenges, such as real-time detection, data scarcity, and generalization to unseen attacks, and deployment under resource-constrained environments typical of vehicular systems. By synthesizing insights from existing works and identifying underexplored areas, this survey aims to guide future research and the development of intelligent, adaptive, and scalable IDS solutions for IoV using generative models.

1.2. Survey Methodology

To identify and review existing studies on Intrusion Detection Systems in the Internet of Vehicles, with a particular emphasis on the use of Generative AI techniques, a comprehensive and methodologically rigorous literature search was conducted. The goal was to capture state-of-the-art approaches, evaluate their performance, and analyze research trends within this emerging domain.

A systematic search of the literature was conducted using Google Scholar. The search aimed to identify relevant studies published between 2015 and 2024, capturing recent trends and advancements in both ML/DL-based IDS and Generative AI-based IDS techniques applied in vehicular networks.

The search was performed using combinations of the following keywords:

("Intrusion Detection System" OR "IDS")

AND ("Internet of Vehicles" OR "IoV" OR "Vehicular Network" OR "V2X" OR "CAN bus") AND ("Machine Learning" OR "Deep Learning" OR "Generative AI" OR "GAN" OR "Transformer" OR "LLM" OR "BERT" OR "GPT").

Titles, abstracts, and keywords were manually reviewed to assess the relevance of each paper. Only studies offering technical depth were considered. To ensure relevance and quality, the following criteria were used to select the surveyed papers:

1.2.1. Inclusion Criteria

- Research applying IDS to vehicular networks, including in-vehicle (IVN) and external communication (V2X) settings.
- Studies involving Machine Learning, Deep Learning, or Generative AI methods such as GANs, Transformers, and LLMs.
- Papers providing sufficient technical detail, including the threat model, dataset, and evaluation metrics.

1.2.2. Exclusion Criteria

- Papers focused on general IoT or cloud environments without specific application to vehicular networks.
- Studies lacking experimental validation or performance metrics.
- Non-English articles or those with limited accessibility or academic rigor.

The inclusion and exclusion criteria were designed to uphold scientific rigor and ensure the selection of relevant, technically sound studies. After applying these criteria, 58 papers were selected, each focusing on both ML/DL- and Generative AI-based IDS within the Internet of Vehicles. These studies encompass a diverse set of approaches, including traditional Machine Learning, Deep Learning, and emerging Generative AI techniques such as GANs, Transformers, LLMs, and hybrid architectures. This categorization provides valuable insight into the current landscape of IDS research in vehicular networks and highlights the growing prominence of generative approaches in this field.

To provide a visual summary, Figure 1 presents a bar chart illustrating the distribution of techniques used across the reviewed studies. The methods are categorized as ML-based, DL-based, GAN-based, Transformer-based, Autoencoder-based (AE), and other Generative AI-driven IDS approaches.

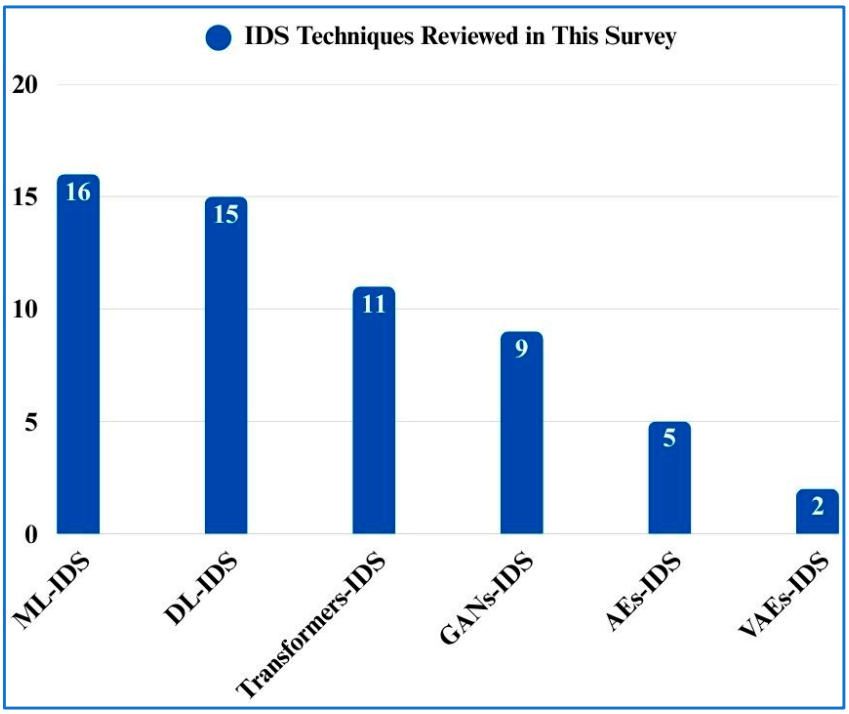


Figure 1. Distribution of IDS techniques reviewed in this survey.

1.3. Survey Organization

This survey aims to explore the potential of Generative AI-based IDS for IoV networks, analyzing existing research, discussing the challenges faced, and suggesting future research directions to improve the security of IoV ecosystems. The paper is structured as follows: Section 2 presents the background, covering IoV communication types, architecture, possible attacks, and a review of classical ML- and DL-based IDS approaches. Section 3 explores core Generative AI techniques for threat detection. Section 4 examines existing Generative AI-based IDS for IoV. Section 5 discusses challenges and future research directions, and Section 6 concludes the study.

2. Background

2.1. IoV Communication Types

The integration of detection systems, control interfaces, and computational tools transforms each automobile in the Internet of Vehicles into an intelligent entity. Each vehicle connects with other elements through a comprehensive communication framework, referred to as V2X. This connectivity aims to promote safer driving by minimizing collisions, providing route guidance, and delivering various information services. Vehicles in this network interact dynamically with all factors influencing their operation [4].

Vehicle-to-Everything (V2X) mainly includes five types of connectivity: Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), Vehicle-to-Pedestrian (V2P), Vehicle-to-Cloud (V2C), and Inter-Vehicle (Inter-V), as illustrated in Figure 2.

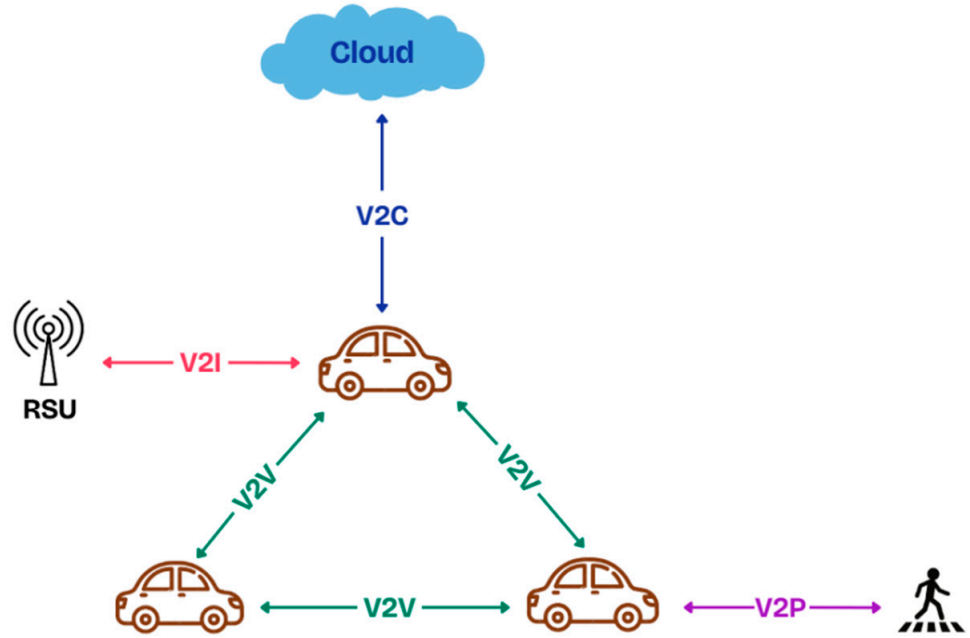


Figure 2. IoV communication types.

Vehicle-to-Vehicle refers to wireless communication between vehicles, where each vehicle can exchange real-time information about speed, location, and direction with other vehicles. Vehicle-to-Infrastructure is the communication established between vehicles and various roadside infrastructure units (RSUs) to enhance transportation efficiency and safety. Vehicle-to-Pedestrian enables vehicles to monitor and interact with pedestrians and cyclists on the road, improving safety and avoiding accidents. Vehicle-to-Cloud involves communication that allows vehicles to connect and exchange data with cloud-based services. Inter-Vehicle (Inter-V) is a communication system that enables the sharing of data from a vehicle's internal sensors among all users inside the vehicle.

2.2. IoV Communication Architecture

There is still no agreement among academics about the structure of the Internet of Vehicles architecture. In Ref. [5], Liu Nanjie mentioned that the IoV architecture closely resembles that of the IoT and defined a three-layer architecture named the "Client-Connection-Cloud" system. The first layer (sensing layer) includes sensors inside the vehicle that collect environmental data and monitor the vehicle's conditions. The network layer is the second layer, which supports different V2X communication modes to ensure seamless connectivity between vehicles, infrastructure, and other devices. The third layer (application layer) integrates statistical tools, data storage, and processing capabilities to enable efficient big data analysis and decision-making, supporting IoV intelligence.

Over the years, several researchers have proposed different IoV architecture designs. Table 2 provides a summary of the existing IoV architectures.

Table 2. Summary of the existing IoV architectures.

Existing Architectures	Layers	Layers Name
2011—Liu Nanjie [5]	Three	1—Client, 2—Connection, 3—Cloud
2016—Kaiwartya et al. [6]	Five	1—Perception, 2—Coordination, 3—Artificial intelligence, 4—Application, 5—Business
2017—Contreras-Castillo et al. [7]	Seven	1—User interaction layer, 2—Data acquisition layer, 3—Data filtering and preprocessing layer, 4—Communication layer, 5—Control and management layer, 6—Business layer, 7—Security layer
2019—KaiLiu et al. [8]	Four	1—Cloud computing, 2—SDN control, 3—Fog computing, 4—Infrastructure
2020—Ji et al. [9]	Four	1—Cloud platform, 2—Edge, 3—Data acquisition, 4—Security authentication
2022—Gao et al. [10]	Three	1—Clients, 2—Endorsement and commitment peers, 3—Ordering services
2023—Wang et al. [11]	Four	1—Terminal, 2—Network, 3—Platform, 4—Service

Considering the evolving security and privacy challenges in IoV systems, the seven-layer IoV architecture proposed by Contreras-Castillo et al. in Ref. [7] stands out as the most suitable IoV architecture for addressing these concerns. Unlike previous designs that lacked integrated security mechanisms, this architecture incorporates a dedicated security layer functioning transversally across all layers to manage authentication, authorization, and transaction accounting, ensuring consistent and system-wide protection. It also features a data filtering and preprocessing layer that ensures only relevant, useful, and non-sensitive data is transmitted, reducing network traffic and minimizing the risk of privacy breaches. Additionally, the model offers an intelligent communication interface that optimizes network selection based on various factors such as network quality, communication requirements, and transaction costs, further enhancing the secure and efficient handling of data.

According to the seven-layer architecture proposed by Contreras-Castillo et al. [7], the IoV system is structured into distinct yet interconnected layers, each playing a specific role in ensuring the system's effectiveness and security. As illustrated in Figure 3, the User Vehicle Interface Layer facilitates communication between the driver and the vehicle, managing alerts and notifications through visual, auditory, or haptic feedback, depending on the context. The Data Acquisition Layer is responsible for gathering data from both internal vehicle sensors and external sources such as other vehicles, traffic infrastructure, and environmental sensors. The Data Filtering and Preprocessing Layer reduces network congestion and data overload by filtering out irrelevant information and prioritizing data based on the vehicle's subscribed services. The Communication Layer selects the most suitable network for data transmission by evaluating parameters such as quality of service,

congestion, security, and cost. The Control and Management Layer coordinates services within the IoV environment, applying traffic policies and inspection mechanisms to optimize information flow. In the Processing Layer, large volumes of data are analyzed either locally or via cloud computing to enhance services and inform infrastructure planning. Finally, the Security Layer acts transversally across all other layers, implementing functions such as authentication, access control, integrity checks, and cyber threat mitigation to ensure system-wide trust and resilience.

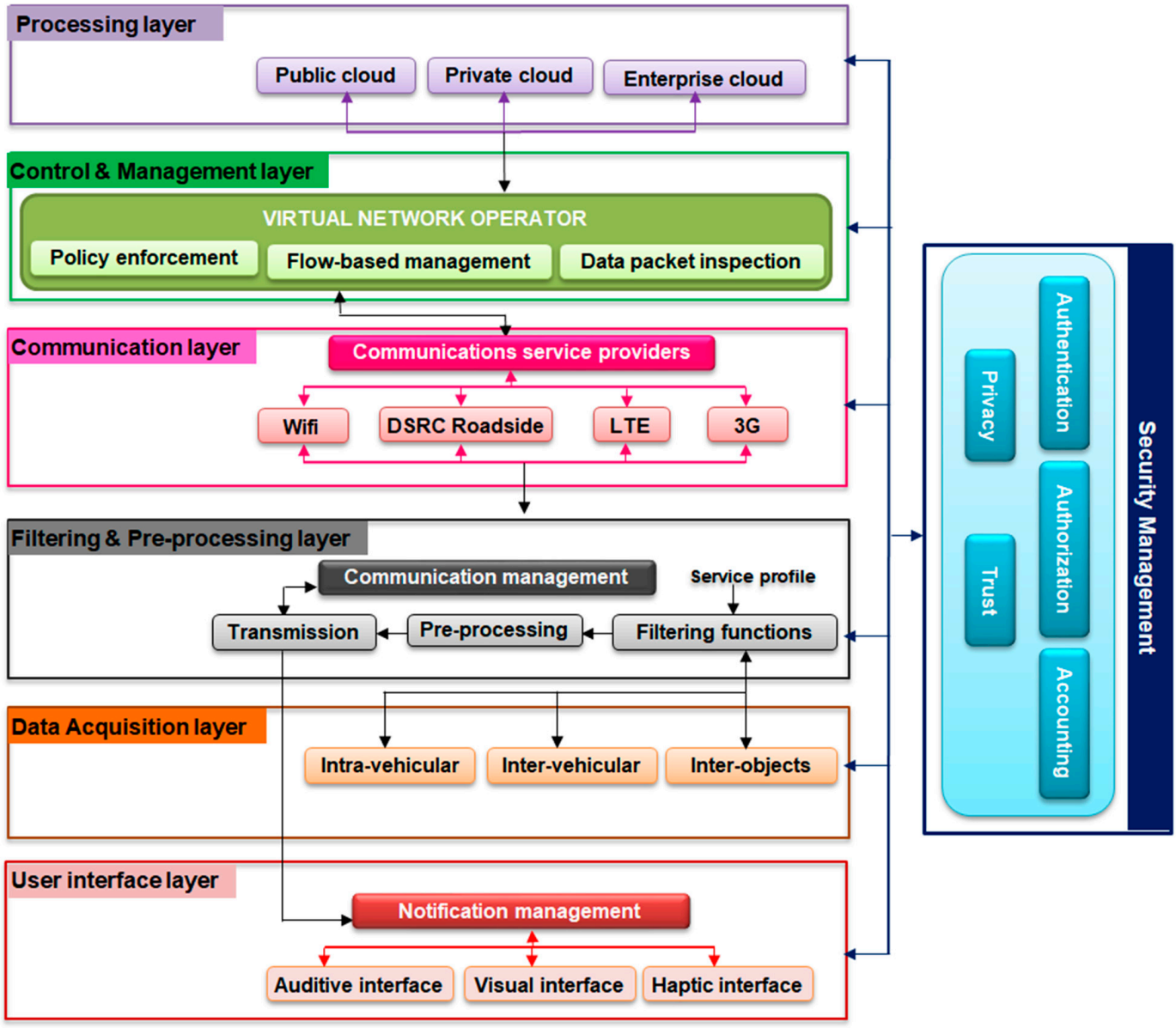


Figure 3. The seven-layer IoV architecture proposed.

2.3. Possible Attacks in IoV

The Internet of Vehicles system is vulnerable to various attacks that can compromise its stability, security robustness, real-time performance, and privacy. These vulnerabilities not only diminish its ability to provide effective services but may also result in severe accidents. This section provides an overview of these attacks in IoV. Table 3 presents a concise summary of the attacks targeting the IoV system, highlighting the attack type, along with the specific security components that are vulnerable, including authentication, availability, and confidentiality.

Table 3. Summary of IoV attacks and their affected security components.

Attack	Attack Type	Affected Security Component
Sybil attack	Active	Authentication, Availability
Masquerading attack	Active	Authentication, Confidentiality
Man-in-the-Middle attack	Active, Passive	Availability, Confidentiality
Replay attack	Passive	Availability
Cookie Theft attack	Active	Authentication, Confidentiality
Message Injection attack	Active	Availability
Message Manipulation attack	Active	Availability
Channel Interference attack	Active, Passive	Availability
Denial of Service (DoS) attack	Active	Availability
Distributed DoS attack	Active	Availability
Eavesdropping attack	Passive	Confidentiality
Message Holding attack	Active	Availability
False Information Flow attack	Active	Availability
Channel Hindrance attack	Active	Availability
Malware attack	Active	Availability
Physical Vehicle damage	Active	Availability
Fuzzy attack	Active	Availability
Guessing attacks	Active	Authentication
Session Linking attack	Active	Authentication, Confidentiality
Black-hole attack	Active	Availability
Forgery attack	Active	Authentication
Wormhole attack	Active	Availability
GPS Deception attack	Active	Confidentiality, Availability

One of the most common attacks in IoV systems is the Sybil attack, in which an attacker creates multiple fake vehicle identities to deceive the network into believing that numerous vehicles are present. This can disrupt the IoV system by causing false traffic information to be reported, leading to incorrect traffic decisions, virtual congestion, and inefficient route planning. Another common attack, the Channel Interference attack, involves degrading the availability of IoV networks by introducing noise or distortions into the communication channels. These attacks are often categorized as jamming attacks [12].

In Cookie Theft attack, the attacker saves cookies and reuses them to gain unauthorized access to network resources whenever needed, posing a serious security risk to IoV systems. Meanwhile, in a Message Manipulation attack, the attacker modifies the content of the messages, resulting in incorrect decisions by the receiving system, which can incapacitate the entire network. In the Channel Hindrance attack, the attacker attempts to disrupt communication channels, slowing or restricting information exchange and impacting real-time applications within the IoV environment [13].

The Denial of Service (DoS) attack overwhelms a resource provider with excessive requests, rendering it unable to handle legitimate requests and severely interrupting vehicular traffic and system operations. In an advanced variant, the Distributed Denial of Service attack, the attacker sends a large volume of traffic from multiple systems to a specific target, causing it to become unavailable and disrupting its normal operations. Another serious

threat is the GPS Deception attack, where the attacker captures and manipulates GPS signals or sends fake signals to mislead a vehicle into thinking it is at a different location than intended. This poses a significant threat to both vehicle and user security [12].

In the False Information Flow attack, authenticated users can be tricked by incorrect or tampered with information, creating a false perception of the IoV environment, leading to wrong decisions that jeopardize the system's security and efficiency. Another attack is the Message Injection attack, where the attacker inserts deceptive data into the IoV network in order to breach the vehicle's security, targeting its electronic control unit or infotainment and telematics systems for unauthorized access.

A Malware attack involves introducing malicious software, such as viruses or worms, into the network to disrupt its functionality. These attacks are difficult to counter due to the decentralized structure of the IoV. A Session Linking attack occurs when an attacker connects two random sessions of a vehicle with other entities in the network, potentially exposing all credentials after minimal calculation [14].

In an Eavesdropping attack, the attacker intercepts and monitors data transmissions, compromising the privacy and confidentiality of the information, while a Message Holding attack occurs when an attacker delays or withholds parts of the message, disrupting the flow of critical information. Physical Vehicle Damage refers to attacks that involve direct harm to a vehicle's components, such as tampering with hardware, damaging sensors, or destroying critical systems, to compromise its functionality or security [15].

The Fuzzy attack occurs when the attacker observes a vehicle's behavior over time to manipulate its patterns and then sends deceptive messages with random or repetitive data to confuse identifiers, disrupting the system's functionality. In Guessing attacks, attackers attempt to gain unauthorized access by guessing a vehicle's identity, passwords, or other authentication credentials, often by intercepting and analyzing transmitted data to extract useful information [13].

In a Replay attack, an attacker repeatedly transmits valid or altered information to a target vehicle, disrupting its real-time functionality and increasing the risk of system malfunction or failure. A Black-hole attack involves an attacker intercepting and absorbing packets of information exchanged between legitimate users, causing the loss of critical data and disrupting real-time communication [16].

The Man-in-the-Middle attack occurs when a malicious vehicle intercepts communication between vehicles (V2V) or between a vehicle and infrastructure (V2I). The attacker may eavesdrop passively to gather sensitive information or actively alter, delay, or delete messages to manipulate outcomes. Another attack is the Forgery attack, where an attacker pretends to be a legitimate user device or onboard unit (OBU), sending fraudulent commands to manipulate or control the network [17].

Another common attack is the Masquerading attack, where the attacker's vehicle falsely uses another vehicle's ID to gain unauthorized access to restricted information. The key difference between this attack and a Sybil attack is that it does not involve creating multiple identities. Lastly, the Wormhole attack occurs when two or more malicious nodes create a false representation of the network by concealing the actual distances between them, forcing legitimate nodes to route data through them, allowing attackers to intercept information and disrupt the network [18].

2.4. Emerging AI-Driven Threats in IoT and Implications for IoV Security

Recent developments in AI-based attack strategies within general IoT systems have revealed a range of sophisticated side-channel and generative techniques that pose significant security risks—many of which carry direct implications for the Internet of Vehicles (IoV). For instance, Cao et al. [19] introduced MagInspector, an unsupervised learning-

based detection mechanism that captures GPU magnetic side-channel patterns to identify cryptojacking behaviors, achieving significant accuracy improvements over supervised baselines. Similarly, MagSign, as introduced in Ref. [20], leverages dynamic magnetic fields generated by IoT screens to perform low-cost user authentication using ambient magnetism, demonstrating the potential of magnetism as both an identifier and a potential vector for spoofing if exploited.

Other research has revealed novel side-channel attacks through ubiquitous IoT hardware. Ni et al. [21] proposed WISERS, which exploits coil whine and magnetic field perturbations during wireless charging to infer user interactions—such as screen unlocking—with over 90% accuracy. Subsequent work, AppListener, as proposed in Ref. [22], showed that radio frequency (RF) energy harvested from Wi-Fi transmissions can be used to detect fine-grained mobile app activity, with classification accuracies exceeding 98%. Additionally, FPLogger, presented in Ref. [23], demonstrated that electromagnetic emissions from in-display fingerprint sensors can be leveraged to reconstruct fingerprint images and conduct successful spoofing attacks on modern smartphones.

Generative models are also playing a dual role in both attacks and defenses. The AccEar framework, described in Ref. [24], employs a conditional GAN to reconstruct unconstrained speech from accelerometer data, bypassing microphone permissions and exposing audio leakage vulnerabilities in voice-based systems. Meanwhile, the F2Key system, proposed in Ref. [25], applies generative modeling to embed user-specific facial features and vocal patterns into a private key, achieving strong resistance against spoofing and replay attacks in wearable devices.

Beyond hardware- and sensor-level threats, Large Language Models (LLMs) introduce further risks. Zhou et al. [26] presented a detailed taxonomy of backdoor attacks in LLMs, highlighting vulnerabilities introduced during training and outlining potential defenses. Complementing this, Wang et al. [27] surveyed the use of LLMs in program analysis, emphasizing their role in automating code comprehension and identifying security flaws—while also introducing new avenues for exploitation if misused.

Although these works are situated in general purpose IoT and mobile contexts, they reveal evolving attack surfaces and capabilities that are increasingly transferable to vehicular systems. Side-channel vulnerabilities, generative model-driven spoofing, and LLM manipulation all exemplify the complexity of future threats. These insights reinforce the need for adaptive, intelligent, and generative IDS frameworks that can address not only classical intrusion patterns but also emerging, multi-modal, and context-aware attack vectors in IoV environments.

2.5. Review of Existing ML and DL Security Approaches in IoV

In this section, we will explore the existing ML and DL security approaches in the Internet of Vehicles, emphasizing their limitations and inefficiencies in addressing emerging security challenges.

2.5.1. Machine Learning-Based IDS

In the Internet of Vehicles, ML-based methods and techniques provided clear advantages in handling large and complex datasets, which led to increased interest from researchers, as highlighted in Table 4, which provides a comparative summary of some ML-based IDSs for the IoV.

Table 4. Comparative summary of ML-based IDS for IoV.

Reference	ML Model	Detection Target	Strengths and Limitations
[28]	KNN, SVM	Classification of intrusion events	✓Fast, effective for structured data ✗Limited by static datasets
[29]	RF, DT, ET, XGBoost	Multiple attack types	✓High accuracy ✗High computational time
[30]	Decision Tree	Network attack detection for vehicles	✓Integration of network and physical features ✗Scalability and flexibility limitations
[31]	DT, RF, KNN	DDoS attacks (UDP-Lag, SYN Flood)	✓High accuracy, effective feature selection ✗Scalability and robustness concerns
[32]	SVM, KNN, C4.5	DDoS attacks	✓Combines multiple classifiers ✗Limited by high computational overhead
[33]	RF, KNN, LightGBM, XGBoost, AdaBoost, SVM, LDA	DDoS attacks	✓High accuracy ✗Dataset-dependent, real-time constraints
[34]	Multiple ML algorithms	Injection attacks on CAN bus	✓Good accuracy and low false alarm rate ✗Missing false negative metrics, high external deployment cost
[35]	Various ML methods (hierarchical)	General vehicular attacks	✓Flexible multi-level detection ✗Lack of performance metrics
[36]	RF, KNN, OCSVM	Spoofing and Jamming attacks	✓Precise attack detection ✗Lacks FPR/FNR and latency metrics
[37]	DT, RF	Flooding attacks in 5G vehicular networks	✓Excellent performance (F1 Score of 1) ✗Narrow attack scope
[38]	KNN, RF, DT, Naïve Bayes	Position falsification in VANETs	✓High accuracy, RSU-based detection ✗Needs dense RSU coverage
[39]	Timing anomaly detection	Injection attacks	✓0% FPR, lightweight ✗Not robust to slow attacks
[40]	Sequence-based anomaly detection	Malicious CAN message injections	✓Low memory and computational footprint, high detection accuracy ✗Limited replay attack detection
[41]	Hamming distance on payloads	Fuzzing, Replay attacks	✓Lightweight, good Fuzzing detection ✗Low detection rate for replay attacks
[42]	Hidden Markov Model (HMM)	Anomalous vehicle states	✓Real deployment. ✗Threshold-dependent
[43]	ML-based physical-layer fingerprinting	ECU and channel identification	✓High accuracy using real signal data ✗Depends on hardware-level variations

Several ML-based IDS have been proposed to strengthen security in the Internet of Vehicles, each offering distinct capabilities and facing specific challenges. Alshammari et al. [28] applied K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) for clustering and classification of intrusion events. Their method demonstrated effectiveness for structured and labeled data, offering fast and interpretable results. However, its performance may decline when faced with more complex or evolving attack patterns. Building on this, Yang et al. [29] proposed more sophisticated ML-based IDS combining traditional classifiers like KNN and SVM with a stacked model of tree-based algorithms, including Random Forest (RF), Decision Tree (DT), Extra Trees (ET), and Extreme Gradient Boosting (XGBoost). To enhance accuracy and mitigate class imbalance, the authors introduced the SMOTE technique. Their system achieved over 99% accuracy, showing strong detection capabilities across multiple attack types, but still struggled to achieve both high accuracy and low latency simultaneously.

Vuong et al. [30] addressed some of these limitations by proposing a decision tree-based IDS that incorporates both network and physical features, acknowledging the unique constraints of vehicular environments, such as mobility and energy consumption. While this integration improves detection in realistic IoV settings, the use of decision trees may restrict scalability and flexibility when dealing with highly complex or large-scale attacks.

Focusing specifically on DDoS threats, Albishi and Abdullah [31] developed a detection model using DT, RF, and KNN to identify UDP-Lag and SYN Flood attacks. Their model demonstrated high accuracy and the benefits of effective feature selection. However, it was limited to only two attack types and lacked scalability and real-world deployment testing. Similarly, Sumathi et al. [32] proposed a hybrid IDS combining SVM, KNN, and the C4.5 classifier. This integration leveraged the strengths of each individual algorithm but suffered from high computational time, echoing the same challenge faced by the stacked model proposed in Ref. [29]. Meanwhile, Halladay et al. [33] utilized the CICDDoS dataset to train binary and multiclass classifiers to detect DDoS attacks. Their study examined eight ML classifiers, including RF, KNN, LightGBM, XGBoost, AdaBoost, SVM, and Linear Discriminant Analysis (LDA). The binary classification scenario achieved an impressive accuracy of approximately 0.99, reflecting the strong potential of ML-based IDS when trained on balanced and comprehensive datasets.

Moreover, Minawi et al. [34] proposed a multi-layered ML-based IDS for CAN bus networks, consisting of threat detection and alerting layers. They evaluated different ML algorithms using a real car hacking dataset with various injection attacks. Although their system achieved good detection accuracy and a low false alarm rate, it lacked details on the false negative rate—a crucial metric. Additionally, placing the IDS externally on the OBD II port could increase costs, which may not be acceptable to vehicle owners or manufacturers. The evaluation of execution time was also insufficient, as it did not reflect the system's true computational performance. Furthermore, Goncalves et al. [35] introduced an Intelligent Hierarchical IDS, dividing the network into four levels with clusters at each, using various ML detection methods. This hierarchical structure allows for flexibility in applying different techniques at each layer. However, although tested on a simulated dataset, the feasibility of this system remains uncertain, as many important evaluation metrics were omitted.

Kosmanos et al. [36] proposed a cross-layer IDS targeting spoofing and jamming attacks in connected vehicle platoons. Their system employs RF, KNN, and One-Class Support Vector Machine (OCSVM), alongside cross-layer data fusion methods. It offers probabilistic outcomes for both known and unknown attacks. Tested in a simulated platoon environment using the Veins simulator, the IDS showed precise detection for both attacks. Nonetheless, its scope was attack-specific, and critical metrics like the false positive rate (FPR), false negative rate (FNR), and detection latency were not addressed. In addition, in the context of 5G-enabled vehicular networks, Sousa et al. [37] presented an IDS model based on DT and RF classifiers to detect Flooding attacks using ns-3-generated data. The Decision Tree classifier performed particularly well, achieving a perfect F1 Score of 1. In a different approach, Sharma and Jaekel [38] proposed an IDS targeting position falsification attacks using a two-consecutive-BSM (2BSM) feature approach, processed at RSUs instead of vehicle OBUs. Evaluated on the VeReMi dataset using KNN, RF, DT, and Naïve Bayes, their system achieved higher accuracy and lower false alarm rates, particularly for difficult position-based attacks. However, it depends on dense RSU coverage, which may limit its practicality in sparse areas.

Song et al. [39] proposed a lightweight IDS based on analyzing time intervals between CAN messages to detect injection attacks. The method leverages the fact that, under normal conditions, Electronic Control Units (ECUs) transmit messages at consistent and predictable intervals. An injection attack disrupts this regularity, typically causing a significant increase

in message frequency—often 20 to 100 times higher than normal. By monitoring these deviations, the system can effectively identify abnormal behavior. Experimental evaluation using real anonymized CAN data from a vehicle demonstrated that the approach achieves high detection accuracy with no false alarms.

Marchetti and Stabili [40] proposed an anomaly-based algorithm designed to detect malicious CAN message injections on the CAN bus of modern vehicles. The algorithm identifies anomalies by analyzing the sequence of message flows and is optimized for low memory and computational requirements, making it suitable for deployment on current ECUs. Its detection performance is demonstrated through experiments on real CAN traffic collected from an unmodified licensed vehicle, showing high detection accuracy and low false positive rates. However, the system struggles to detect replay attacks, achieving a detection rate of only 20 to 40%.

Dario et al. [41] introduced an intrusion detection algorithm for CAN bus systems that leverages the Hamming distance between successive payloads from different ID classes. Designed with minimal memory requirements, the method is well suited for deployment on modern vehicle ECUs. The approach was tested using traffic data collected from a legally operated, unmodified vehicle. Experimental results indicated near-perfect detection rates for message injection attacks involving fuzzing in both NoRange and SmallRange categories. However, its effectiveness significantly dropped for MidRange attacks, with detection rates between 20 and 30%, highlighting its limitations in identifying replay-based threats.

Narayanan et al. [42] proposed the use of a Hidden Markov Model (HMM) to identify abnormal vehicle behaviors. Real-world data was collected from legally operated vehicles, and the anomaly detection task was framed as a classification problem. The CAN bus data was processed as sequential observations using a sliding window technique, with each observation represented by a feature vector (e.g., speed, RPM, door status). The HMM was trained to model normal behavior and compute the posterior probabilities of observed sequences. If an observation sequence falls below a defined probability threshold, it is flagged as anomalous, indicating a potential unsafe state.

Avatefipour et al. [43] proposed a machine learning-based fingerprinting approach that links CAN messages to their originating sources by exploiting unique physical-layer artifacts caused by hardware imperfections. The method extracts a feature vector of 11 statistical attributes from both the time and frequency domains and uses these to train a neural network classifier. Evaluated on data from 16 CAN channels and four identical ECUs sending the same messages, the system achieved detection accuracies of 95.2% for channel classification and 98.3% for ECU identification.

2.5.2. Deep Learning-Based IDS

Deep Learning techniques have gained significant attention in the development of IDS for IoV due to their ability to automatically learn hierarchical features from raw data, offering substantial advantages over traditional machine learning methods. This has encouraged researchers to explore and implement DL-based IDSs, as illustrated in Table 5, which compares some DL-based IDSs for IoV.

Table 5. Adaptive comparative summary of DL-based IDS for IoV.

DL Model	Reference	Detection Target	Strengths and Limitations
MLP	[44]	DoS attack, U2R attack, R2L attack, Probe attack	✓High accuracy with simple design ✗Needs deep architecture for better performance
DNN	[45]	General malicious packet injection attacks	✓High accuracy and real-time detection ✗Time-consuming training
	[46]	CAN bus signal anomalies	✓High detection accuracy, robust to external intrusion attacks ✗Cannot detect attacks originating from internal ECUs
	[47]	CAN malicious attacks	✓Tested on real data from 3 vehicles ✗Complexity in embedded setups
DBN	[48]	VANET intrusions	✓Good for classification, mitigates overfitting ✗Rarely used independently, slower
CNN	[49]	Real-time intrusion detection at vehicle terminals	✓Lightweight, real-time detection ✗Limited handling of temporal data
	[50]	CAN bus intrusion	✓Effective online/offline detection ✗Requires external CAN devices
DCNN	[51]	DDoS attack	✓High accuracy, good for RSU traffic ✗Latency not addressed
LSTM	[52]	Fuzzy, RPM, DOS, GEAR attacks	✓Good for sequential data, multiple intrusion types ✗Resource-intensive
	[53]	CAN bus attacks	✓High accuracy ✗High latency (128 ms), large overhead
Bi-LSTM + DNN + Bloom Filter	[54]	Internal and external IoV attacks	✓Short training, low detection latency ✗Lacks FPR and FNR metrics
CNN, LSTM	[55]	Wormhole, Sybil, DoS, Infiltrating Transfer, DDoS, Black-hole, Brute Force attacks	✓Efficient, high detection performance ✗Model complexity
CNN-LSTM, MLP, LSTM, CNN	[56]	Disruptive, Replay, DoS, Sybil, Hybrid	✓High accuracy, low latency, MEC-friendly ✗Struggles with similar attack types
HTM	[57]	CAN anomaly detection	✓Outperforms RNN and HMM in precision and recall ✗Relies on synthetic attack data
RNN-LSTM with cloud offloading	[58]	Cyberattacks	✓Outperforms traditional classifiers in accuracy ✗Limited to robotic testbed

A variety of DL models have been widely applied to IoV intrusion detection, specifically, Artificial Neural Networks, especially the Multi-Layer Perceptron, which have shown considerable effectiveness in this area. For instance, Anzer et al. [44] developed a Distributed Intrusion Detection System based on MLP to monitor malicious traffic within IoV networks, and it successfully identified a range of attacks. Although the system achieved high accuracy with a small feature set, it still faced the limitation that deeper networks are required for improved results.

Building on the concept of ANN, Deep Neural Networks add multiple hidden layers, allowing the model to extract higher-level features from raw data. In Ref. [45], an advanced IDS based on DNN is presented, designed to enhance the security of in-vehicle networks, specifically targeting CAN buses. While the system demonstrates high accuracy and real-time performance, its main limitation is the time-consuming training phase. In related efforts, Xun et al. [46] proposed an IDS that monitors message transmission on the CAN bus using voltage signals unique to each ECU, extracted through statistical features and a DNN model. It achieved high detection accuracy, robust to external attacks, but cannot detect attacks originating from vehicle ECUs.

Zhang et al. [47] proposed a two-stage intrusion detection system combining rule-based logic and a deep neural network (DNN) for real-time attack detection on the CAN bus. The first stage uses lightweight, rule-based techniques to rapidly identify violations in message periodicity and regularity, reducing the computational load on the second stage. The second stage, based on a DNN, is designed to detect attacks missed by the initial filter. The system was evaluated using real CAN traffic collected from three different vehicles—a Honda Accord, an Asia-brand vehicle, and a US-brand vehicle. Results showed high effectiveness, with detection rates between 99.91% and 99.97% and per-message detection times ranging from 0.53 ms to 0.61 ms.

On the other hand, Vitalkar et al. [48] applied a Deep Belief Network model—a notable extension of DNNs—to intrusion detection within Vehicular Ad Hoc Networks, demonstrating the effectiveness of DBN for both binary and multiclass classification tasks. However, DBNs are more often used as tools for understanding the principles and capabilities of DL rather than for direct application in intrusion detection systems.

In the field of Convolutional Neural Networks, which are particularly effective for spatial data processing, researchers have explored their potential for IoV intrusion detection. Peng et al. [49] designed lightweight CNN-based IDS for vehicle terminals, where they represented network traffic as images to differentiate between normal and abnormal traffic patterns in real time. This approach highlights the ability of CNNs to handle spatial data, which is valuable in dynamic and complex IoV environments. Desta et al. [50] introduced a CNN-based IDS for CAN bus systems, tested on real car data. While achieving reliable detection both online and offline, it relied on external CAN devices, increasing system costs and limiting vehicle compatibility, as well as requiring full retraining when the data distribution changes.

An extension of CNN, called Deep Convolutional Neural Networks, introduces additional layers and parameters to better handle large-scale datasets. Nie et al. [51] applied DCNN to design a data-driven IDS for detecting abnormal traffic caused by intrusion events, achieving better performance in terms of accuracy, although it did not fully address the latency concerns inherent in such deep architectures.

Recurrent Neural Networks, especially Long Short-Term Memory networks, are highly effective for sequential data, such as time series data in IoV networks. Ashraf et al. [52] designed an LSTM-based autoencoding algorithm to detect multiple intrusion types in IoV networks. Another approach by Agrawal et al. [53] proposed a DL-based IDS, using LSTM models combined with thresholding and error reconstruction, for CAN bus intrusion detection, achieving high accuracy but at the cost of large computational overhead and a 128 ms latency while neglecting other important performance metrics.

In a hybrid approach, Khan et al. [54] used a combination of a bloom filter and DNN-based bidirectional LSTM architecture to detect both internal and external attacks in IoVs. The system demonstrated a short training period and very low detection latency, making it highly efficient. However, it lacked comprehensive metrics such as False Positive Rate (FPR) and False Negative Rate (FNR). Furthermore, Zeng et al. [55] combined LSTM and CNN models to create a hybrid IDS capable of detecting intrusion events passing through On-Board Units (OBUs). Their system achieved higher detection performance while requiring fewer resources. This research emphasizes the potential of RNN-based techniques for IoV, where large volumes of continuous data are generated, underscoring the need for efficient time series modeling in intrusion detection. Moreover, Alladi et al. [56] introduced a MEC-based IDS for IoV using four DL models: LSTM, CNN-LSTM, MLP, and CNN. The CNN-based sequence-image classifier outperformed others in accuracy and latency, making it ideal for real-time detection on edge devices like Raspberry Pi. However, it struggled slightly with similar attack types like disruptive and replay attacks.

Wang et al. [57] introduced a distributed intrusion detection system based on Hierarchical Temporal Memory (HTM) to detect modification and replay attacks on the CAN bus. HTM models are capable of real-time prediction by learning the temporal patterns of data sequences. This system operates in an online learning mode, continuously updating its memory as new data flows in, allowing it to adapt to changes without requiring prior knowledge of the CAN protocol structure. By monitoring ongoing data streams, the IDS can effectively detect both known and previously unseen attacks on the vehicle network. Loukas et al. [58] introduced an RNN-LSTM-based intrusion detection system for robotic vehicles, utilizing deep learning to capture the temporal patterns of cyberattacks. To address the high computational demands, a cloud-based offloading framework was implemented. The study also explored the trade-off between offloading and detection latency.

While machine learning and deep learning approaches have made significant strides in addressing security challenges in the Internet of Vehicles, they still face substantial limitations, such as high computational costs, latency issues, scalability challenges, and the risk of high false positive rates. These constraints highlight the growing need for Generative AI-based IDS, which present more flexible, resource-efficient, and adaptive solutions capable of addressing the dynamic and increasingly sophisticated security threats faced by IoV networks. By leveraging generative models, IDS systems can enhance threat detection accuracy and adapt in real time to evolving attack strategies.

3. Generative AI for Threat Detection

Generative Artificial Intelligence refers to a branch of artificial intelligence that focuses on creating new, realistic content, such as text, images, or programming code, based on input data or prompts. Unlike traditional AI systems that are primarily used for tasks like classification or prediction, Generative AI leverages deep generative models to learn the underlying patterns and distributions within data [59]. Additionally, Generative AI is designed to understand and replicate complex patterns within data, enabling systems to simulate human-like creativity, reasoning, and problem-solving. These systems learn from vast datasets to model underlying structures and relationships, allowing them to generate outputs that not only resemble real-world data but also enhance decision-making, simulate environments, and even fill in missing information.

This versatility makes Generative AI a powerful tool across various domains, such as the Internet of Vehicles field, where it can be a valuable asset in designing advanced Intrusion Detection Systems that effectively identify and mitigate malicious attacks, ensuring the security and reliability of vehicular communication networks.

Some of the most prominent Generative AI techniques include GANs, Transformers, VAEs, Generative Diffusion Models (GDMs), and AutoRegressive Models (ARMs). Each of these methods employs a unique approach. For example, GANs consist of two neural networks—the generator and the discriminator—that work in opposition to improve the quality of the generated data. Transformers have revolutionized natural language processing by efficiently handling long-range dependencies in data sequences. VAEs focus on learning a probabilistic mapping of data into a latent space, which allows for controlled generation of new samples. GDMs work by progressively denoising data, while ARMs predict the next element in a sequence based on previous ones.

These techniques form the foundation of Generative AI and act as building blocks for developing complex models capable of producing realistic and meaningful outputs. Figure 4 below illustrates these techniques, and the following sections will delve deeper into each one.

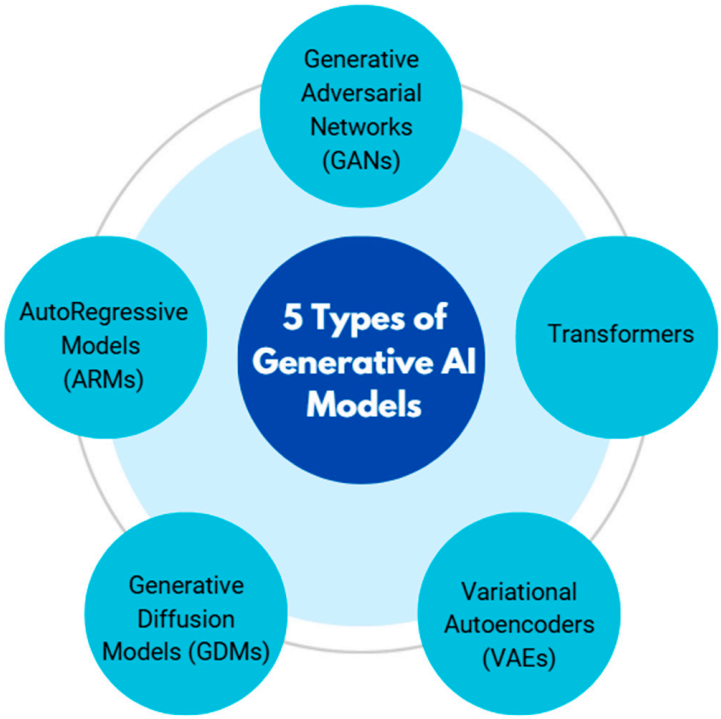


Figure 4. Types of Generative AI models.

3.1. Generative Adversarial Networks (GANs)

Generative Adversarial Networks are a type of DL-based generative model introduced by Ian Goodfellow in his 2014 paper [60]. The core idea behind GANs is to train two competing neural networks—a generator and a discriminator—in a game–theoretic framework. The generator creates fake data, while the discriminator's task is to distinguish between real and fake data. Over time, the generator improves its ability to generate realistic samples, and the discriminator becomes better at identifying fakes, leading to a dynamic optimization process that ultimately results in high-quality generated data.

The architecture of GANs consists of two primary components: the generator and the discriminator. The generator takes a random noise vector as input and transforms it into synthetic data, while the discriminator evaluates whether the input data is real or fake. As shown in Figure 5, the generator (G) learns to create data similar to the real data (X_data), and the discriminator (D) classifies inputs as real or fake. Both networks are trained simultaneously, with the generator aiming to minimize the discriminator's ability to distinguish between real and generated data, while the discriminator seeks to improve its accuracy in identifying fakes. This adversarial process continues until the generator produces realistic data that the discriminator can no longer reliably distinguish from real data [61].

To provide a deeper understanding of the internal mechanics of GANs, this training can be conceptualized as a minimax game between the two networks, each with opposing goals. Initially, the generator produces poor-quality synthetic data from random noise. The discriminator evaluates both real and generated samples, generating feedback that guides the generator's updates. With each iteration, the generator adjusts its parameters to produce more realistic data, while the discriminator concurrently becomes better at detecting fakes. Over time, this competitive interaction leads to a dynamic equilibrium in which the generator produces data distributions that closely mimic real ones, and the discriminator's ability to distinguish real from fake becomes effectively neutralized. This adversarial learning paradigm underlies the GAN's strength in capturing and modeling complex data distributions with high accuracy.

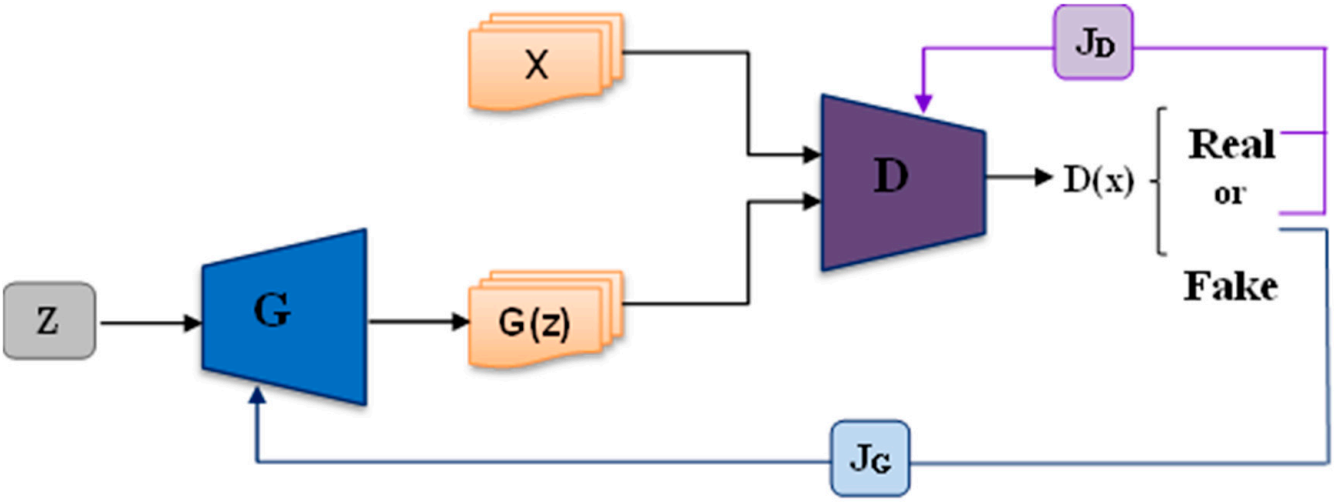


Figure 5. The architecture of the GAN.

In the field of security, GANs can be applied to detect attacks by generating synthetic attack data that mimics real-world attack scenarios. By training on both real and fake attack data, the discriminator can improve its ability to identify malicious behavior in a system. GANs could be particularly useful in generating attack signatures for intrusion detection systems, enabling them to recognize and respond to new and previously unseen attacks. This capability can enhance the detection of sophisticated threats, such as zero-day attacks, by providing a broader range of attack samples for training the security models.

3.2. Transformers

Transformers are a type of DL architecture introduced by Vaswani et al. in 2017 in the groundbreaking paper "Attention Is All You Need" [62]. Unlike earlier models that relied heavily on recurrence (like RNNs) or convolution (like CNNs), Transformers process sequential data entirely through self-attention mechanisms. They were originally developed for machine translation but have since become the foundation for many state-of-the-art models in natural language processing (NLP), such as BERT, GPT, and T5. The transformer model is designed to handle variable-length input sequences and to capture long-range dependencies with greater efficiency than traditional sequence models.

In the original paper [62], the transformer architecture is composed of an encoder and a decoder, each consisting of multiple identical layers—typically six in the base model. The encoder receives the input sequence and transforms it into a series of high-dimensional representations. The decoder, in turn, takes these representations and generates the output sequence, one element at a time, using both the previous output and the encoded input.

Each layer in both the encoder and decoder contains two core components: a multi-head self-attention mechanism and a feed-forward neural network. The attention mechanism allows the model to weigh the importance of different tokens in the input sequence when computing each token's output representation. Multi-head attention, in particular, enables the model to focus on different positions in the sequence simultaneously, which helps capture a wide variety of dependencies and relationships. The output of these layers is passed through residual connections, followed by layer normalization to stabilize and speed up training.

A key innovation in Transformers is the use of positional encodings to provide a sense of order to the input tokens, since the model does not inherently understand sequence order without recurrence. These encodings are added to the input embeddings and use sinusoidal functions that allow the model to learn and extrapolate positions within sequences.

The Transformer architecture can be seen in Figure 6, where it is visually split into two halves—the encoder on the left and the decoder on the right. Each encoder layer contains a self-attention mechanism followed by a feed-forward network. Each decoder layer adds a third sub-layer for encoder–decoder attention, allowing the decoder to attend to the encoder's output. The overall structure is highly modular and parallelizable, which contributes significantly to faster training times compared to RNN-based models. This design achieves state-of-the-art results in tasks like machine translation while substantially reducing computational cost.

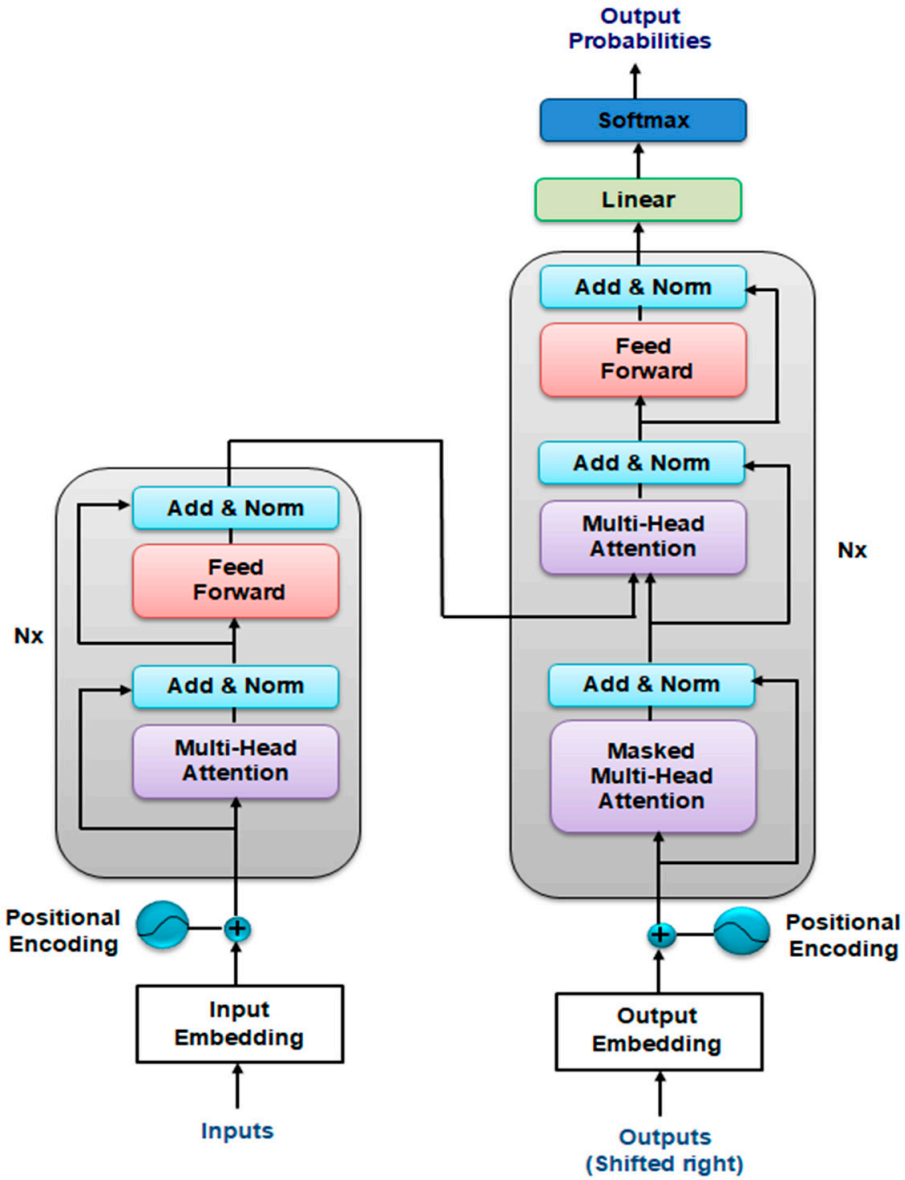


Figure 6. The architecture of the Transformer.

In simple terms, the Transformer is a model that learns how words (or tokens) in a sentence relate to each other, without reading them one by one like traditional models. Instead, it looks at all words at once and decides which ones are important to focus on for understanding meaning. For example, when translating a sentence or detecting an anomaly, it can connect distant words or values that are related—even if they are far apart in the sequence. Its architecture enables it to process data faster and more accurately, making it especially powerful for language tasks and time series analysis in security applications.

In the context of cybersecurity, Transformers have demonstrated significant potential, particularly in detecting various types of attacks. Their ability to capture global context and complex patterns makes them ideal for analyzing sequential data such as system logs, network traffic, or command sequences, which are common in security-related tasks. Unlike traditional rule-based or statistical models that struggle with evolving threats, Transformers can learn from large datasets and generalize to unseen attack vectors. For example, in Intrusion Detection Systems, Transformers are used to analyze sequences of network events or system logs to detect anomalies linked to attacks. They are also employed in detecting phishing attempts, analyzing malware behaviors, and inspecting binary code for embedded threats. Furthermore, Transformers' interpretability through attention weights provides valuable insights into which parts of the data contributed to a detection decision, supporting more transparent and explainable cybersecurity solutions.

3.3. Variational Autoencoders (VAEs)

Variational Autoencoders, introduced by Kingma et al. [63], are a class of deep generative models designed to learn latent representations of data in a probabilistic framework. Unlike classical Autoencoders that map input data to a fixed code, VAEs encode each input into a distribution over the latent space, typically parameterized as a Gaussian. This allows for efficient sampling and generation of new data points. The learning objective of a VAE is derived from variational inference, where the model aims to maximize the likelihood of the data while regularizing the latent space to follow a known prior distribution. This is accomplished by optimizing the Evidence Lower Bound (ELBO), which balances reconstruction quality with the closeness of the learned latent distribution to the prior [64]. VAEs have been widely adopted for tasks such as image synthesis, anomaly detection, and unsupervised representation learning due to their flexibility and strong mathematical foundations.

In terms of architecture, VAEs are composed of two neural networks: an encoder and a decoder. The encoder receives the input data and outputs the parameters (mean and variance) of a latent Gaussian distribution. A latent vector is then sampled from this distribution using the reparameterization trick, which allows gradients to flow through the sampling operation during backpropagation. The decoder then attempts to reconstruct the original input from the sampled latent vector. This process forces the model to learn compact, informative latent representations. As illustrated in Figure 7 of the referenced paper [64], the VAE architecture includes a stochastic component where random noise sampled from a standard normal distribution is transformed using the encoder's outputs to generate the latent variable. This sampled variable is passed to the decoder for reconstruction. The overall structure allows for both efficient encoding of inputs and flexible generation of new data by sampling directly from the latent space.

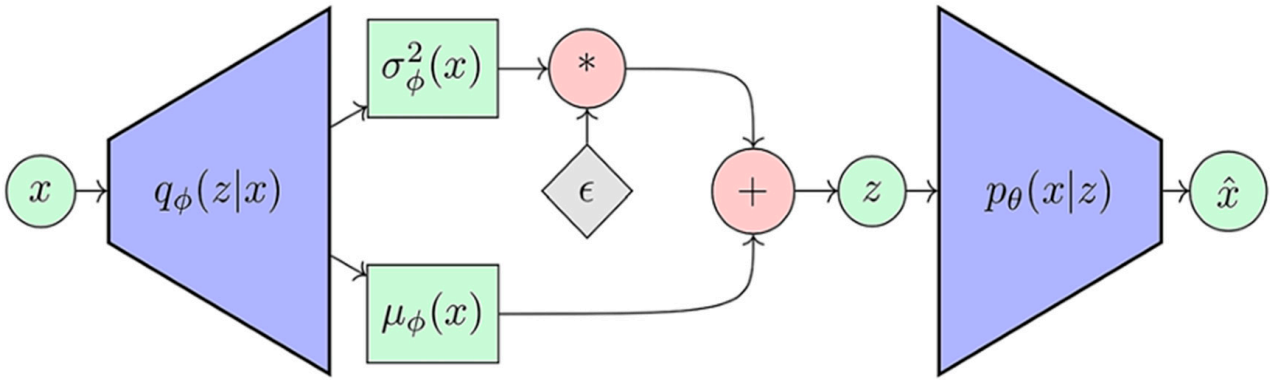


Figure 7. The architecture of the VAE [64].

VAE can be thought of as a smart compressor and generator. It first compresses input data (like images or messages) into a smaller abstract form (called a latent representation), but instead of producing a single fixed output, it learns a range of possible representations using probability. This allows the model not only to recreate the original data but also to generate new, similar data by sampling from that learned range. This probabilistic nature makes VAEs particularly effective for detecting unusual patterns, such as anomalies in security systems, or generating realistic synthetic data for training purposes.

VAEs also offer powerful capabilities in the field of cybersecurity, especially for attack detection in complex and dynamic environments. Since VAEs are generative models trained to learn the normal distribution of data, they can be employed to detect anomalies that deviate from expected behavior. In security applications such as intrusion detection systems, VAE can be trained on benign network traffic or system logs. During deployment, any input that significantly deviates from the learned distribution—indicated by a high reconstruction error or a low likelihood under the latent space—can be flagged as potentially malicious. This makes VAEs highly effective for detecting novel attacks, zero-day exploits, or subtle deviations in behavior that rule-based systems may overlook. Furthermore, the flexibility of VAEs allows them to be adapted to various types of data, including packet flows, system events, or even sensor data in IoV environments. By capturing the underlying structure of normal behavior, VAEs provide a robust and unsupervised approach to identifying cybersecurity threats without requiring labeled attack data.

3.4. Generative Diffusion Models (GDMs)

Generative Diffusion Models are a class of probabilistic generative models that learn to model complex data distributions by simulating a two-step process inspired by non-equilibrium thermodynamics. First introduced by Sohl-Dickstein et al. in Ref. [65], GDMs involve gradually adding noise to data through a forward diffusion process until the data becomes pure noise and then learning a reverse diffusion process that denoises this noise step by step to reconstruct the original data. This framework allows for both high model flexibility and tractability, offering exact sampling and likelihood evaluation, which are significant advantages over traditional generative models.

The architecture of GDMs is based on two main components: a forward diffusion process and a reverse generative process. The forward process incrementally adds noise to a data sample through a Markov chain, transforming it into a simple, known distribution such as Gaussian noise. The reverse process is then learned to invert this transformation, reconstructing data from noise. This is achieved by estimating small perturbations needed to reverse each diffusion step. The entire model becomes a sequence of denoising steps that, together, generate data samples [65]. A visual representation of this structure and workflow is shown in Figure 8.

To better understand their operation, GDMs can be interpreted as models that learn to reverse a structured noise-injection process. Initially, data is gradually corrupted by noise until it becomes indistinguishable from pure randomness. The model is then trained to denoise this input incrementally, step by step, in order to reconstruct the original data. This forward–reverse dynamic enables GDMs to generate high-quality samples that closely follow the distribution of real data. Owing to their systematic design, GDMs are particularly effective at modeling intricate data patterns and producing diverse, realistic outputs.

GDMs have emerged as a valuable approach for enhancing intrusion detection systems, particularly in identifying malicious activities or anomalies in network traffic. Their ability to model complex, high-dimensional data distributions makes them well suited for learning the normal behavior of network or system data. Once trained on clean (attack-free) data, the reverse diffusion process can be used to detect discrepancies between the generated

(expected) data and real-time observations. If the input data significantly deviates from the learned distribution, it may indicate an attack or anomaly. Moreover, GDMs' probabilistic nature allows for the quantification of uncertainty, enabling robust detection even in the presence of noise or adversarial manipulations. As a result, GDMs serve as a powerful tool for modern security applications, including detecting novel or stealthy cyberattacks in dynamic and complex environments.

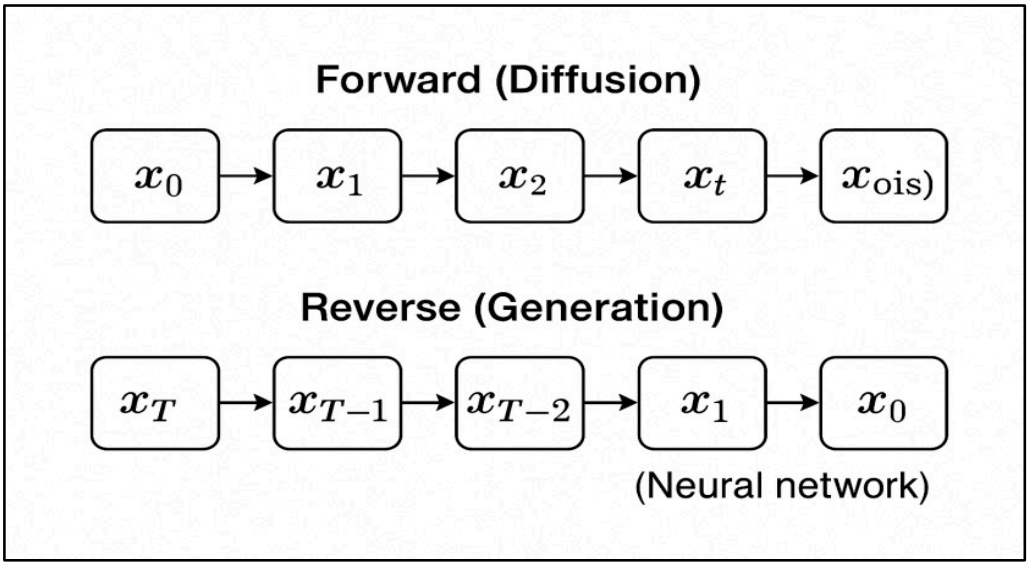


Figure 8. The architecture of the GDM.

3.5. AutoRegressive Models (ARMs)

AutoRegressive Models are a class of statistical models used to analyze and predict time series data by regressing current values on their own previous values (lags). According to the paper [66], ARMs are extended to account for potential regime shifts through Markov Switching, resulting in models where the autoregressive parameters vary depending on an unobserved state or "regime" that follows a Markov process. These models capture both time dependency and structural changes in data, which are common in economic and financial systems. The general formulation presented includes the possibility of an infinite number of past observations influencing the current value, forming what is known as a General AutoRegressive Model with Markov Switching (GARMS) [66].

In ARMs, the current value Yt depends on past values and a random shock. In a Markov-switching setting, the model becomes:

$$\mathrm{Yt} = \mu(\mathrm{Xt}) + \sum \beta\mathrm{i}(\mathrm{Xt})\mathrm{Yt} - \mathrm{i} + \sigma(\mathrm{Xt})\mathrm{nt}$$

where Xt is the hidden regime at time t, and the coefficients μ, βi, σ vary with Xt. The regime follows a Markov chain, and estimation is conducted via maximum likelihood, often using matrix operations to manage the hidden states. This setup allows the model to adapt to changes in behavior across time. This means the model can switch between different patterns depending on the current regime, making it flexible for real-world time series that experience shifts over time [66].

To clarify their function, AutoRegressive Models (ARMs) can be understood as models that predict the present based on the past. They accomplish this by linking current values in a time series to previous values—capturing how past behavior influences future outcomes. When extended with Markov Switching, the model adapts to changes in patterns over time by assuming that the system operates under different hidden "regimes" or states. Each

regime follows its own rules, and the model switches between them based on a probabilistic process. This allows ARMs to handle shifts in behavior, making them well suited for analyzing complex, time-dependent data that may change across different conditions.

AutoRegressive Models can be used in cybersecurity to detect attacks by modeling normal system behavior over time, such as network traffic or login patterns. By predicting future values based on past data, ARMs can flag anomalies when actual behavior deviates from expected trends. These deviations may indicate attacks like intrusions or DDoS events. Using Markov-switching, ARMs enhances detection by allowing the model to adapt to different normal states, improving accuracy and reducing false alarms. This makes ARMs a powerful tool for early threat detection in dynamic environments.

Table 6 presents a comparative summary of the discussed generative models, outlining their core mechanisms and main technical strengths.

Table 6. Comparative summary of Generative AI models.

Generative Model	Core Mechanism	Key Strengths
GANs	Involves adversarial training between two networks: a generator (produces synthetic data) and a discriminator (evaluates authenticity).	—High-quality data synthesis —Learns complex data distributions
Transformers	Utilizes self-attention mechanisms to process all elements of an input sequence in parallel, capturing long-range dependencies efficiently.	—Handles long-range dependencies —Scalable and parallelizable —Pretrained on large data
VAEs	Learns a probabilistic latent representation by encoding and decoding input data, optimizing a balance between reconstruction and generalization.	—Probabilistic reasoning —Low reconstruction error alerts —Good latent representations
GDMs	Learns to reverse a structured noise process by progressively denoising the corrupted input, based on diffusion and generative modeling.	—High sample quality —Robust to noise —Good uncertainty estimation
ARMs	Predicts future values in a sequence based on past observations; can be extended with regime-switching mechanisms.	—Effective in time series —Adapts to changes over time —Interpretable

4. Generative AI-Based IDS for IoV

As connected vehicles become increasingly vulnerable to cyber threats, researchers have turned to Generative AI-based Intrusion Detection Systems to enhance security in vehicular networks. Numerous studies have explored a variety of techniques to detect and mitigate attacks, leveraging models such as GANs, Transformers, VAEs, GDMs, ARMs, and others to improve detection accuracy and adaptability. Table 7 presents a comparison of various Generative AI-based IDS approaches for cybersecurity in connected vehicles, highlighting the core techniques employed in each.

In particular, GANs have proven effective and suitable for detecting both known and unknown threats in real-time vehicular environments. Several works have investigated this direction, including Khalil et al. [67], who introduced a BiGAN-based IDS that captures latent features from network traffic, allowing the system to detect both known and unknown attacks in IoV systems. The BiGAN architecture enhances anomaly detection through bidirectional encoding and decoding processes. Using the CICIDS2017 dataset, this model demonstrates robust intrusion detection performance, but it demands high computational resources. In contrast, Poongodi et al. [68] integrated federated learning with a multi-level distributed GAN model, offering a solution tailored for resource-constrained environments in IoV. This decentralized learning approach reduces the burden on central servers and

improves scalability and efficiency. However, it introduces synchronization overhead and communication complexity, especially in dynamic vehicular environments.

Moreover, Shu et al. [69] proposed a collaborative intrusion detection framework utilizing a multi-discriminator GAN. This enables multiple distributed SDN controllers to jointly train an intrusion detection model for an entire VANET without directly exchanging sub-network flows. The framework was validated using KDD99 and NSL-KDD datasets on an emulated cloud server with three distributed Software-defined Networking (SDN) controllers. Although the authors rigorously validated their system in both IID and non-IID contexts with mathematical proofs, no numerical results for evaluated metrics were presented. Furthermore, Ref. [70] proposed an anomaly detection algorithm using GANs to address the imbalance between normal and attack traffic in vehicular networks. The algorithm, trained solely on normal traffic data, employs the Gini index for dimensionality reduction and normalization functions to enhance training speed. Moreover, in Ref. [71], the authors used GANs to identify driving anomalies, leveraging multiple modalities such as CAN bus data, physiological signals, and environmental information. While this method significantly enhances the detection of irregular driving patterns, its scalability is limited, and it cannot easily incorporate additional modalities.

Table 7. Comparison of Generative AI-based IDS for IoV.

GAI Technique	Ref	Main Approach	Detection Target	Network Focus	Performance Summary	Strengths and Limitations
GANs	[67]	BiGAN (Encoder–Discriminator)	Known and unknown attacks	V2V	Acc: 92.15% F1: 0.961 Latency: 83.71 s FPR: 7.3 FNR: 0.6	✓Robust and strong performance ✗High computational resources
	[68]	Distributed GANs + Federated Learning	Malicious traffic behavior	V2X	Acc: 98.92% F1: 0.77	✓Scalable and decentralized ✗Communication and synchronization complexity
	[69]	Collaborative Multi-Discriminator GAN (SDN)	Intrusion detection	V2X	---	✓Collaborative framework, validated on datasets ✗No numerical evaluation results
	[70]	GAN-based Anomaly Detection	Traffic anomalies	V2X	Acc: 94.9%	✓Addresses traffic imbalance ✗Trained only on normal data
	[71]	Multi-modal GANs	Driving anomaly detection	V2X	---	✓High detection accuracy ✗Poor scalability for new modalities
	[72]	GAN-based IDS	CAN messages	IVN	Acc: 98% Latency: 0.18 s	✓High accuracy ✗Limited to known CAN formats and lacks generalization
	[73]	Dual-discriminator GAN with ID encoding	CAN-FD bus	IVN	Acc: 99.93% Latency: 0.15 ms per message	✓High accuracy (99.93%) ✗Specialized for ID-based detection only
	[74]	Combining CAN communication matrix with image-based GANs	Tampering detection	IVN	---	✓High precision and accuracy ✗Requires complex preprocessing
	[75]	ACGAN with two-stage classification	Attacks classification	IVN	F1: 99.23 Latency: 0.203 ms (multicore) 0.538 ms (single-core) FPR: 0.36 FNR: 0.96	✓High accuracy and low computational cost ✗Design complexity

Table 7. *Cont.*

GAI Technique	Ref	Main Approach	Detection Target	Network Focus	Performance Summary	Strengths and Limitations
Transformers	[76]	Transformer-based attention network + transfer learning	Multi-class attack detection	IVN	Acc: 1-ER = 99.94% F1: 1.00 (single-message input) Latency: 11.6 ms/batch	✓Good classification, adaptable ✗CAN-specific, scalability limits
	[77]	MFVT	Anomaly traffic detection	V2X	Acc: 99.96% (IDS 2012) 99.99% (IDS 2017) F1: 0.9995 (IDS 2012) 1.00 (IDS 2017) FPR: 0.000175 (IDS 2012)	✓Captures diverse input aspects ✗Limited in long-range dependencies
	[78]	CAN-Former (self-supervised)	Anomalies detection	IVN	Acc: 99.75% (Kia Soul), 99.48% (Chevrolet Spark), 98.35% (Sonata) F1: 0.9873 (Kia Soul), 0.9892 (Chevrolet Spark), 0.9702 (Sonata)	✓No handcrafted features ✗Evaluation limited to specific datasets
	[79]	FL-EC with FSFormer	Intrusion detection	V2X	---	✓High accuracy, improves privacy ✗Computational edge overhead
	[80]	Fine-tuned encoder- based Transformer	Misbehavior via BSM sequences	V2X	Acc: 99.66% F1: 99.66% FPR: 0.34	✓Good detection, adaptable ✗Computational challenges
	[81]	LLM (Mistral-7B)	Misbehavior detection	V2X	Acc: 98% F1: 0.98	✓High detection accuracy and efficient use of resources ✗Resources constraints, privacy issues
	[82]	SecureBERT, BERT, LLAMA-2	CAN intrusion detection	IVN	Acc: 99.99% F1: 0.999993 Latency: 14 messages/s FAR: 3.1×10^{-6}	✓Effective, better with large models ✗High computational cost
	[83]	CANBERT	Multiple CAN attacks	IVN	Acc: 100% F1: 1.00	✓High precision and accuracy ✗High computational requirements
	[84]	BERT with self-attention	Harmful texts detection	V2X	Acc: 96.65%.	✓High accuracy (96.65%) ✗Text-only focus, needs labeled data, and high resource demands
	[85]	IoV-BERT-IDS	Network intrusion detection	IVN + V2X	---	✓Improved detection accuracy ✗Requires heavy fine-tuning
	[86]	BERT + masked training	CAN message injection attacks	IVN	F1: 0.81 to 0.99 Latency: 0.8 to 3.8 ms Model size: 20–70 MB ParametersL: ~2.9 M–3.2 M	✓Outperforms PCA/AEs ✗Focused on CAN-ID sequences
VAEs	[87]	CVAE + RF	Traffic anomaly detection	Network Traffic	F1: 0.85 to 0.99 Latency: 82.72 s (MAWILab-2018) 46.37 s (ISCX-2012)	✓Improved feature learning, good accuracy ✗Varies across datasets and attack types
	[88]	A-VAE + CNN + Bi-LSTM + Attention	Traffic video anomaly detection	V2X	EER: 27% (Crossroad1), 31% (Pedestrian) Latency: 0.0149 sec/frame on Crossroad 0.0192 sec/frame on Pedestrian	✓High real-time accuracy, unsupervised learning ✗Needs diverse training data

Table 7. *Cont.*

GAI Technique	Ref	Main Approach	Detection Target	Network Focus	Performance Summary	Strengths and Limitations
AEs	[89]	AE + MLP	DDoS attacks	V2X	Acc: 98.34% F1: 98.18%	✓High accuracy and F1 score ✗Limited to DDoS
	[90]	Multiple LSTM-AEs	CAN bus anomalies	IVN	Acc: 97% F1: 0.98 Latency: 0.5 ms FPR: 0.01	✓Good accuracy with dynamic data ✗Limited to CAN protocol
	[91]	Conv-AE + GAN (semi-supervised)	General intrusions	IVN	F1: 99.84% Latency: 0.63 ms (GPU), 0.69 ms (CPU)	✓Strong detection accuracy ✗High training complexity
	[92]	Denoising AE + Attention Mechanism	Real-time anomalies	IVN	---	✓Robust and generalizable ✗Compared only to traditional ML methods

Acc—Accuracy; F1—F1-score; Lat—Latency; FPR—False Positive Rate; FNR—False Negative Rate; FAR—False Alarm Rate; EER—Equal Error Rate. "---" indicates that the metric was not specified in the referenced study.

As we shift focus to in-vehicle networks (IVNs), Seo et al. [72] proposed a GAN-based Intrusion Detection System, a model that uses GAN to detect both known and unknown attacks on in-vehicle CAN bus networks. By employing a dual-discriminator approach, GIDS identifies attacks using only normal data, achieving high accuracy in detecting known attacks and around 98% accuracy for unknown ones. However, it focused only on known CAN traffic formats and was limited by the quality of generated data. In a similar vein, Wang et al. [73] focused on the CAN-FD protocol and proposed a GAN with dual discriminators to detect ID-based intrusions. By encoding message IDs into binary images and feeding them into the GAN, the system distinguishes between normal and malicious ID patterns. The model achieves remarkable accuracy (99.93%) and low latency but is limited in its application to CAN-FD ID segments.

In another approach, Xie et al. [74] introduced an innovative model that combines a CAN communication matrix with image-based GANs to detect tampering in CAN signals. This method reduces false positives by categorizing messages into different transmission types and constructing images using consecutive messages from the same ID. However, it requires complex preprocessing and is dependent on ID-specific analysis. Further, Zhao et al. [75] proposed a two-stage IDS using an Auxiliary Classifier GAN (ACGAN), where the first stage classifies known attacks into categories, and the second stage handles out-of-distribution (unknown) samples. This approach demonstrated high classification accuracy with minimal computational cost, making it suitable for real-time use in embedded systems.

Another Generative AI technique is Transformers, which have shown significant promise in intrusion detection for IoV, making several researchers adopt them to build intelligent IDS models to enhance detection performance and adaptability.

To begin with, Ref. [76] presents a transformer-based attention network as an innovative IDS for CAN buses in vehicles. Driven by the urgent need to address CAN protocol vulnerabilities, this method leverages self-attention mechanisms for efficient multi-class classification of attacks, such as replay attacks. Furthermore, it employs transfer learning to improve detection performance, especially on small datasets from different car models. In a similar vein, Ref. [77] introduces the MFVT model, an anomaly traffic detection system combining a feature fusion network with a vision transformer architecture. Although it does not explicitly mention Generative AI, this approach significantly improves intrusion detection by accurately identifying suspicious network behaviors, thereby safeguarding network integrity and reducing potential economic losses. Notably, the transformer encoder's multi-head attention mechanism allows the model to focus on different parts of the input

data concurrently. However, it sometimes struggles to capture long-range dependencies due to limitations in the self-attention mechanism.

Likewise, Ref. [78] proposes CAN-Former IDS, a Transformer-based model that predicts anomalies in CAN communications, considering both ID sequences and message payload values. This system stands out by employing fully self-supervised training, eliminating the need for handcrafted features and demonstrating its capabilities using a survival analysis dataset collected from three vehicles. Moving forward, Ref. [79] introduces a federated learning-based edge computing (FL-EC) architecture to enhance the privacy and security of V2X communications. This study also presents FSFormer, which integrates a feature attention mechanism to dynamically select critical features for IDS tasks. Experiments using the UNSW-NB15 and CSE-CIC-IDS2018 datasets reveal that FSFormer achieves superior accuracy and F1 scores, outperforming several baseline deep learning models.

In addition to these, Ref. [80] applies Transformers for misbehavior detection in V2X communications. By fine-tuning an encoder-based transformer on normal Basic Safety Message (BSM) sequences, the method detects misbehavior by comparing transformed sequences against predefined thresholds. While the results are promising, processing long BSM sequences (length 200) may present computational challenges. Another notable contribution is [81], which proposes a Large Language Model (LLM)-empowered Misbehavior Detection System (MDS) using the Mistral-7B model, fine-tuned at the edge for real-time detection and paired with a larger cloud-based LLM for in-depth analysis. The experiments, conducted on the extended VeReMi dataset, show that Mistral-7B outperforms other LLMs, achieving 98% accuracy in misbehavior detection. The system processes vehicular data into time windows, transforming them into textual prompts for LLM inference. The main advantage of the system is its high detection accuracy and efficient use of resources, especially with Mistral-7B's performance outperforming other LLMs like LLAMA2-7B and RoBERTa. However, challenges include resource constraints in real-time detection and concerns about data privacy when transferring data to the cloud. Additionally, Ref. [82] explores the application of well-known LLMs, including SecureBERT, BERT, and LLAMA-2, for CAN intrusion detection. Their study confirms the effectiveness of these models, highlighting the superior performance of larger models like LLAMA-2 in capturing complex message distributions.

In addition to the aforementioned works, BERT-based models have gained significant attention in IDS due to their ability to model semantic structures and relationships in data. For instance, the research in [83] introduces CANBERT, a language-based intrusion detection model for the CAN bus, leveraging transformer models to detect malicious attacks. It provides a comprehensive analysis using a CAN dataset derived from a realistic driving scenario, including normal and malicious data from various attack types, such as Denial of Service (DoS), fuzzy, and impersonation attacks. The approach achieves high precision and accuracy in detecting all attacks. However, it has high computational requirements and processing time for pretraining, making real-time training on resource-constrained in-vehicle networks challenging. Similarly, another approach [84] adapts a BERT-based model with self-attention (BERT-SA) for enhancing cybersecurity in vehicular networks. This model is designed to detect phishing, violent, and extremist text content, achieving high accuracy (up to 96.65%) when tested with sentiment analysis datasets. While effective, the model faces limitations, such as focusing solely on text data, relying on labeled datasets, and requiring high computational resources.

Furthermore, Ref. [85] introduces IoV-BERT-IDS, a hybrid Network IDS (NIDS) for both in-vehicle and extra-vehicle networks. Using a semantic extractor to preprocess network traffic for BERT, this model benefits from BERT's bidirectional contextual feature extraction, significantly improving detection accuracy in IoV scenarios. Meanwhile,

Ref. [86] employs a sliding window of CAN-IDs fed into BERT, using a masked language model to predict candidate CAN IDs. Abnormalities are detected when actual IDs fall outside the predicted set. This BERT-based approach demonstrates superiority over traditional ML methods like PCA and Autoencoders in anomaly detection.

VAEs are generative models used in IoV systems to detect anomalies, and they have been applied in various works. For instance, the authors in Ref. [87] propose a method that combines a Conditional VAE (CVAE) with a Random Forest classifier to address issues of data generalization and overfitting in network traffic anomaly detection. The CVAE introduces traffic packet labels into the latent space, allowing better feature learning, while RF enhances classification accuracy. Although this method improves detection rates within a specific dataset, its performance can vary across different datasets and attack types. Meanwhile, in Ref. [88], an Attention-based VAE (A-VAE) is introduced for anomaly detection in traffic video streams. The framework integrates 2D-CNN, Bi-LSTM layers, and an attention mechanism to effectively capture spatial and temporal features. Trained in an unsupervised manner on regular video sequences, this model achieves high real-time detection accuracy in dynamic V2X environments. However, its unsupervised nature means performance might drop if the training data does not adequately represent the diversity of normal traffic patterns.

Autoencoders (AEs) are widely used in intrusion detection systems due to their ability to identify anomalies based on reconstruction errors. To enhance detection accuracy in vehicular networks, many studies have incorporated additional techniques into AE-based models. For instance, Wei et al. [89] proposed a hybrid deep learning-based model to detect DDoS attacks by combining an AE and an MLP network while using the CICDDoS2019 dataset to train the model. In this case, the Autoencoder performs feature selection, and the Multi-layer Perceptron classifies the attacks. Impressively, the model achieved an accuracy score of 98.34% and an F1 score of 98.18%, indicating its effectiveness for DDoS detection. Adding to the diversity of approaches, Kim et al. [90] introduced a novel method for detecting anomalies in in-vehicle networks through the use of multiple LSTM-Autoencoder models. This approach was developed in response to the lack of robust security measures within the CAN protocol, which remains vulnerable to a variety of attacks. The proposed method, enhanced by GAI (Genetic Algorithm-based Initialization), delivers a solution that is both accurate and effective for vehicular applications by considering various dynamic features such as transmission intervals and payload value changes.

Moreover, Hoang and Kim [91] proposed a semi-supervised IDS that integrates a convolutional Autoencoder with a GAN. Their framework involves three training phases: initial Autoencoder training on unlabeled data, joint training with a GAN discriminator, and final fine-tuning using labeled data—achieving strong accuracy with minimal supervision. Further advancing this area, Wei et al. [92] developed AMAEID, an anomaly detection method combining a denoising Autoencoder with an attention mechanism. By injecting Gaussian noise and emphasizing key input features, their model improved robustness and generalization, though it was evaluated only against traditional ML baselines.

4.1. Standardized Evaluation and Benchmarking Challenges

Despite promising results, comparing Generative AI-based IDS models across studies remains challenging due to the absence of standardized evaluation protocols. Many surveyed papers report different combinations of metrics, often calculated on distinct datasets or under varied experimental setups. In several instances, key performance indicators such as latency, computational overhead, or false negative rates are omitted entirely, making direct comparison difficult and potentially misleading.

Furthermore, the use of diverse datasets introduces additional variability. These datasets differ in terms of complexity, attack types, feature richness, and their representativeness of real-world vehicular environments—all of which significantly influence the reported performance of IDS models.

To ensure fair and reproducible evaluation, the establishment of standardized benchmarking frameworks is essential within the domain of generative IDS for the Internet of Vehicles. Future research efforts should adopt unified evaluation protocols that include a core set of metrics reported consistently across studies. In addition, the development of publicly available vehicular datasets that reflect a wide range of real-world scenarios and attack types is crucial. Open-source implementations of IDS models and experimental setups should also be encouraged to promote transparency and support comparative analysis.

Adopting such standardized practices would significantly improve the reliability of performance evaluations, facilitate meaningful cross-study comparisons, and accelerate the translation of generative IDS technologies from experimental research to real-world automotive applications.

4.2. Comparative Analysis: Generative AI's Role in Addressing Traditional ML/DL Limitations

While ML- and DL-based intrusion detection systems have played a foundational role in enhancing vehicular cybersecurity, their effectiveness remains constrained by several critical limitations, including high false positive rates, poor generalization to unseen threats, reliance on extensive labeled datasets, and inadequate adaptability in dynamic environments. Generative AI-based IDS, as demonstrated across recent works, provide compelling solutions to these challenges by leveraging advanced architectures capable of modeling complex data distributions, generating synthetic training data, and learning contextual patterns more effectively.

One of the most prominent shortcomings of conventional ML/DL methods is their inability to reliably detect zero-day attacks or generalize to unfamiliar inputs. Most models are trained on predefined, labeled datasets and often fail when confronted with novel attack patterns. In contrast, generative models such as the GAN-based IDS proposed by Seo et al. [72] detect unknown CAN bus attacks by being trained solely on normal traffic, achieving approximately 98% detection accuracy for unseen threats. Likewise, transformer-based models like CAN-Former [78] and IoV-BERT-IDS [85] rely on self-supervised or masked training schemes to learn long-range dependencies and temporal structures within CAN or V2X sequences, which significantly boosts their capacity to recognize anomalous behavior beyond the scope of their training data. Furthermore, the LLM-empowered IDS [81], by using Mistral-7B, outperformed other LLMs in misbehavior detection and achieved 98% accuracy using real V2X data.

Generative models also address the common issue of class imbalance in cybersecurity datasets by generating realistic synthetic attack samples. This has been exemplified in models such as BiGAN [67], ACGAN [75], and the traffic-imbalanced GAN detector in Ref. [70], where synthetic augmentation helped mitigate data sparsity and improved detection robustness. These approaches allow the IDS to train effectively even when the available malicious data is limited or imbalanced, a condition under which most traditional classifiers tend to underperform or overfit.

In addition, generative models exhibit superior performance in minimizing false positives and false negatives by more accurately capturing the underlying structure of normal system behavior. The GAN model proposed in Ref. [73], which utilized dual discriminators and ID encoding for CAN-FD traffic, achieved 99.93% accuracy with low latency, showcasing its capacity for fine-grained anomaly detection. Similarly, the A-VAE model from [88] and attention-enhanced denoising Autoencoders in Ref. [93] succeeded in

distinguishing irregular behavior from benign anomalies, which are often misclassified in ML-based IDS due to rigid decision boundaries or handcrafted feature dependencies.

Sequential data modeling is another domain where generative approaches outperform conventional DL architectures. Transformer-based IDS models such as FSFormer [79], CANBERT [83], and SecureBERT [82] use attention mechanisms and contextual embedding to model message semantics and relationships across time, allowing for precise anomaly localization. Unlike CNNs or MLPs, which primarily extract spatial or static features, these generative Transformers are adept at processing long-term dependencies and extracting higher-order semantics necessary for recognizing subtle or evolving attack behaviors.

Moreover, the deployment feasibility of generative IDS has improved significantly. Federated transformer architectures like the one proposed in Ref. [79] enable edge-level detection with privacy preservation by avoiding central data collection. Similarly, lightweight GAN and Autoencoder models such as those in Refs. [75,93] are increasingly optimized for real-time inference under the computational constraints typical of vehicular onboard units.

Altogether, the reviewed literature demonstrates that Generative AI-based IDS are not only technically more capable but also practically more scalable and adaptive than traditional ML and DL approaches. By learning from limited data, adapting to emerging threats, and reducing false alarms while maintaining efficiency, these models establish a more resilient and intelligent defense framework for securing the Internet of Vehicles against a continuously evolving spectrum of cyber threats.

4.3. Contextual Suitability of Generative Models

The performance of generative models in Intrusion Detection Systems for the Internet of Vehicles greatly depends on the specific operational constraints and architectural settings of the vehicular environment. In this section, we examine the contextual suitability of three widely used generative approaches—GANs, Transformers, and VAEs—by evaluating their alignment with four critical IoV requirements: real-time latency, robustness under network variability, hardware compatibility, and memory/computation constraints.

In latency-sensitive scenarios such as V2V and V2I communications, real-time detection is essential to avoid cascading system failures or collisions. Transformers, owing to their parallel processing and efficient attention mechanisms, are well suited for such tasks, especially when optimized. VAEs, while not designed for real-time classification per se, offer low-latency inference via reconstruction error and are typically faster than GANs. In contrast, GANs—due to their dual-network architecture and adversarial training—are generally unsuitable for strict real-time applications unless significantly simplified.

When it comes to robustness under varying network conditions, Transformers excel because they can learn contextual dependencies across long input sequences and adapt to changing input patterns. VAEs are also resilient, particularly in unsupervised settings where they can flag deviations from the learned normal distribution. GANs, however, often require high-quality, balanced data to maintain stability; their performance can deteriorate in noisy, heterogeneous vehicular environments.

From a hardware compatibility standpoint, VAEs are the most efficient and deployable, given their relatively lightweight structure and low computational footprint. Transformers, especially when compressed or fine-tuned, can also operate on edge devices, although their initial model size may require optimization. GANs are the most demanding and often need offloading to more powerful computing platforms unless aggressively pruned or simplified.

Regarding memory and computational constraints, VAEs again offer a practical solution for in-vehicle or roadside deployment due to their minimal resource requirements. Transformers fall in the moderate range but are increasingly viable due to edge-friendly

variants. GANs are the least favorable, requiring high memory and prolonged training time to maintain stable generator–discriminator dynamics.

A summary of this contextual analysis is provided in Table 8, highlighting the strengths and limitations of each generative model type across the four identified criteria.

Table 8. Comparative evaluation of GANs, Transformers, and VAEs based on key IoV criteria.

Criteria	GANs	Transformers	VAEs
Real-Time Latency	✗ Poor (slow training and inference)	✓ Good (fast inference with attention)	✓ Good (fast reconstruction-based scoring)
Robustness Under Network Variability	✗ Sensitive to noise and imbalance	✓ High (context-aware sequence modeling)	✓ Moderate (probabilistic generalization)
Hardware Compatibility	✗ Low (GPU required)	△ Moderate (requires optimization)	✓ High (lightweight and deployable)
Memory and Computation Constraints	✗ High (dual models, unstable training)	△ Moderate–High (manageable with pruning)	✓ Low (suitable for edge devices)

Although these models show promise, each comes with specific limitations. GANs frequently suffer from mode collapse, where the generator outputs limited variation, reducing the ability of IDS to detect diverse threats. Several studies, including [70,71], report that GANs perform poorly with imbalanced or low-quality data, potentially generating unrealistic or repetitive patterns and failing to generalize to new attacks. Transformer-based models, such as CAN-Former [78] and IoV-BERT-IDS [85], demonstrate high accuracy due to their attention mechanisms and ability to model long-range dependencies. However, studies such as [81,83] acknowledge that large-scale models like Mistral-7B and LLAMA-2 deliver superior performance but come at the cost of extensive memory consumption, long pretraining times, and limited real-time viability on vehicle-grade hardware. Even with edge-optimized versions, deployment remains challenging in constrained environments. VAEs, while lightweight and well-suited for edge devices, have their own vulnerabilities. As demonstrated In Refs. [87,88], VAEs may suffer from latent space collapse, where different inputs are mapped to nearly identical latent vectors, degrading the model's ability to distinguish between benign and malicious behaviors. These limitations can lead to high false negatives, especially when trained on low-diversity or poorly normalized datasets. The unsupervised nature of many VAE-based models further complicates detection in real-world traffic where anomalous patterns are subtle and not well represented in training data.

Choosing the right generative model for an IoV IDS requires balancing detection performance against resource availability. GANs, while resource-intensive, typically achieve high detection accuracy, especially for rare or previously unseen attacks. They are excellent in training-phase augmentation but may not be viable for real-time, low-power applications without extensive optimization. Transformers strike a balanced trade-off. When pretrained and fine-tuned appropriately, they provide both high accuracy and adaptability to evolving threats. However, their training and deployment still require moderate computational resources, which may limit applicability in ultra-constrained environments unless paired with edge-optimized architectures. VAEs, though often yielding slightly lower detection precision compared to GANs and Transformers, offer superior efficiency and deployability. Their ability to generalize from normal data makes them highly effective in environments with limited labeled datasets and constrained hardware.

In conclusion, each generative model brings a distinct set of advantages and limitations. GANs excel in accuracy and data generation; Transformers offer robust, context-aware modeling; and VAEs shine in low-resource, edge-compatible scenarios. The optimal choice

depends on the specific operational context, threat model, and system constraints of the targeted IoV application.

5. Challenges and Future Research Directions

5.1. Challenges

Generative AI-based IDSs for the Internet of Vehicles face several pressing challenges that hinder their practical deployment and effectiveness. A primary concern is the high computational and memory requirements of generative models, such as GANs and VAEs. These models demand extensive processing power and are not well suited to the resource-constrained nature of vehicular edge devices. In addition, their longer training and inference times pose difficulties for real-time intrusion detection, where split-second responses are often necessary to ensure passenger safety and vehicle functionality.

Moreover, the absence of vehicle-grade hardware acceleration tailored to support complex deep learning architectures further limits their deployment. Most in-vehicle processors lack dedicated AI acceleration units (e.g., NPUs or GPUs), making it impractical to run generative models locally without major architectural redesigns.

Another major challenge is accurately distinguishing between malicious behavior and benign system anomalies. In IoV environments, abnormal data patterns may stem from non-malicious causes such as sensor drift, hardware failures, or environmental factors. However, generative models often struggle to differentiate these from actual cyberattacks, resulting in false positives that reduce trust in IDS alerts. This challenge is intensified by privacy concerns, as many AI models rely on centralized data collection and processing, which risk exposing sensitive vehicle or user information when data is transmitted to cloud servers.

This problem is compounded by V2X communication bandwidth limitations, which can delay or restrict the transmission of real-time IDS outputs or model updates. Such constraints are especially critical in dense urban environments or high-speed traffic scenarios, where latency directly impacts the system's ability to respond to threats in time.

Further complicating the issue is the limited availability of high-quality, real-world datasets for training and evaluating IDS. Most existing models are built and tested using simulated or synthetic data, which may not capture the full range of threats and behaviors encountered in real driving scenarios. Only a few studies, such as [39,43,47], have validated their methods using data collected from actual vehicular systems, including real CAN traffic or physical testbeds. However, these deployments are generally small scale and focus on conventional ML-based IDS, not generative models. As such, a clear gap persists between academic research and large-scale, production-grade integration of Generative AI into vehicle systems. This highlights the urgent need for future research to progress beyond simulation and lab validation toward industry-aligned pilot deployments.

In addition, there is a scarcity of publicly available datasets that capture adverse driving conditions such as weather-induced sensor noise, nighttime scenarios, or rare failure events, which limits the generalizability of IDS models across diverse operational contexts.

Additionally, the absence of standard evaluation benchmarks and large-scale real-world testing platforms makes it difficult to assess how these systems will perform under dynamic conditions, such as mixed traffic environments involving both autonomous and human-driven vehicles.

Integration with existing vehicle architectures presents another bottleneck. Modern vehicles are governed by legacy ECUs and domain-specific network protocols, which often lack the flexibility to accommodate the high-bandwidth data pipelines and low-latency inference engines needed for Generative AI. Seamless incorporation of IDS components

without disrupting functional safety or system determinism remains a significant engineering hurdle.

Real-world deployment of Generative AI models in safety-critical automotive systems is also constrained by the absence of clear certification pathways. Compliance with standards such as ISO/SAE 21,434 [93] and UN Regulation No. 155 (WP.29) requires transparency, robustness, and traceability—criteria that black-box generative models (e.g., GANs, VAEs) often fail to meet. Until generative IDS solutions can demonstrate certifiable behavior and integrate seamlessly into existing automotive functional safety frameworks, regulatory acceptance will remain a significant bottleneck.

Finally, the evolving nature of cyber threats presents another significant obstacle. Many current IDS systems are static in design, lacking the ability to learn and adapt to new attack methods over time. As attackers continuously develop novel techniques, IDS that do not incorporate mechanisms for ongoing learning and adjustment risk becoming obsolete or ineffective. This limitation highlights the need for more flexible, intelligent systems that can evolve alongside the threat landscape.

5.2. Adversarial Threats to Generative Models

Despite the promising capabilities of Generative AI models in enhancing intrusion detection systems within IoV, these models themselves are vulnerable to various adversarial threats that can compromise their reliability and security. One prominent risk involves GAN poisoning and training data manipulation, where attackers inject maliciously crafted or mislabeled samples into the training dataset. This contamination can cause the generator to produce misleading attack signatures or confuse the discriminator, ultimately degrading the IDS' detection performance. In federated learning contexts, where data is distributed and often not centrally verified, poisoning attacks become even harder to detect and mitigate.

Another significant threat category is input tampering and evasion attacks, in which adversaries introduce subtle perturbations to input data designed to mislead the IDS. These adversarial examples are often imperceptible to humans but can result in a complete misclassification by generative models such as VAEs- or Transformer-based detectors. For instance, a manipulated CAN message with minimal alterations might evade detection if the model overly relies on learned patterns rather than robust features. This raises concerns about the deployment of such models in real-time, safety-critical vehicular environments.

Additionally, model stealing and reverse engineering attacks pose risks to the intellectual property and functional integrity of generative IDS. In scenarios where models are deployed at the edge (e.g., within OBUs or RSUs), adversaries can query the model extensively to replicate its behavior, recover its architecture, or even expose sensitive training data. Once replicated, stolen models can be studied offline to generate adversarial traffic or optimize evasion strategies. These threats necessitate the integration of robust defensive mechanisms, such as adversarial training, watermarking, differential privacy, and model access limitations, in generative IDS design.

5.3. Ethical and Privacy Considerations

As Generative AI becomes integral to securing IoV ecosystems, several ethical and privacy challenges arise that must be explicitly addressed to ensure responsible deployment. A primary ethical concern is the potential misuse of synthetic attack data. While Generative Adversarial Networks (GANs) and other generative models can produce valuable synthetic datasets to train or evaluate IDSs, such data could be exploited by malicious actors if leaked or improperly shared. The availability of realistic synthetic attack patterns may inadvertently lower the barrier to crafting new, evasive cyberattacks targeting vehicular networks.

Furthermore, data privacy concerns are particularly pronounced in federated learning scenarios, where generative models are trained across distributed vehicle nodes without centralized data collection. Although federated learning is privacy-preserving by design, recent studies have shown that generative models can memorize and leak sensitive user data through side channels or model updates. Attacks like membership inference or gradient inversion can extract identifiable information from model parameters, undermining the foundational privacy guarantees of such systems. To mitigate these risks, techniques such as differential privacy, secure multi-party computation, and encrypted model aggregation should be incorporated into federated generative IDS designs.

Equally important is the issue of transparency and explainability, especially given the safety-critical nature of vehicular systems. Many Generative AI models—including Transformers and VAEs—operate as black boxes, providing limited insight into why a particular decision was made. In high-stakes contexts such as autonomous driving or real-time traffic coordination, explainability is essential to enable human oversight, legal accountability, and system debugging. Incorporating interpretable AI techniques—such as attention visualization, saliency maps, or post hoc explanation frameworks like SHAP or LIME—can help make generative IDSs more transparent and trustworthy to stakeholders.

5.4. Future Research Directions

To overcome these challenges, future research must focus on making Generative AI-based IDS more practical, efficient, and adaptable for real-world IoV environments. One important direction is the development of lightweight, resource-efficient generative models that can be deployed at the edge without compromising detection accuracy. Techniques such as model compression, quantization, and knowledge distillation can significantly reduce the computational load, making it feasible to run IDS directly on vehicle control units or roadside infrastructure.

In parallel, privacy-preserving ML methods should be adopted to address data confidentiality concerns. Federated learning, for example, allows vehicles to collaboratively train models without sharing raw data, thereby protecting sensitive information while still benefiting from collective learning. Incorporating mechanisms like differential privacy or encrypted model updates can further strengthen security and trust in data handling.

Improving the quality and diversity of training data is another crucial area. Researchers should invest in creating large-scale, realistic datasets that include a variety of driving conditions, traffic densities, and attacker behaviors. These datasets should reflect the complexity of real-world scenarios to improve model generalizability and robustness. Furthermore, integrating data from multiple modalities—such as LiDAR, cameras, GPS, and V2X communication—can help enhance the system's ability to detect context-specific threats and reduce false positives.

The development of standardized evaluation frameworks and real-world testbeds is also essential. Conducting long-term field trials in diverse vehicular settings will help validate IDS performance and reliability under realistic operational conditions. These platforms can serve as benchmarks for comparing different approaches and guiding future improvements.

Lastly, future IDS should be designed with adaptability in mind. By incorporating continual learning, meta-learning, and online training techniques, systems can dynamically evolve in response to new threat patterns and changing network conditions. Additionally, the integration of technologies like blockchain for secure logging and software-defined networking (SDN) for flexible traffic management can further enhance the scalability, transparency, and effectiveness of IDS in complex IoV infrastructures.

6. Conclusions

The research field addressed in this survey is the development of Intrusion Detection Systems for IoV, with a particular focus on the integration of Generative AI techniques. As IoV networks become more interconnected and susceptible to sophisticated cyberattacks, traditional IDS approaches fall short in detecting evolving threats. This survey highlights the significant potential of generative models—such as GANs, VAEs, and Transformers—in enhancing IDS performance by generating synthetic training data, improving detection accuracy, and adapting to previously unseen attack patterns. The main contributions of this work include presenting the first comprehensive review dedicated to Generative AI-based IDS in IoV, offering a structured taxonomy that classifies these models by architecture, deployment context, and attack type, and critically analyzing their effectiveness in addressing real-time detection, data scarcity, and deployment constraints. By bridging existing gaps in the literature, this survey aims to guide future research toward building intelligent, scalable, and resilient IDS solutions tailored to the unique security demands of the IoV ecosystem.

Author Contributions: I.M. was involved in conceptualization, collection of data, writing the paper, and designing the figures. D.E.B., S.A., H.T.-C. and F.I.C.-P. supervised the content and structure of the paper and contributed to collection of data, editing and revisions. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors gratefully acknowledge the support of the Autonomous University of the State of Quintana Roo (UQROO) in this publication.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

IDS	Intrusion Detection System
IoV	Internet of Vehicles
GAI	Generative Artificial Intelligence
GAN	Generative Adversarial Network
VAE	Variational Autoencoder
GDM	Generative Diffusion Model
ARM	AutoRegressive Model
ML	Machine Learning
DL	Deep Learning
IVN	In-Vehicle Network
CAN	Controller Area Network
VANET	Vehicular Ad Hoc Network
IoT	Internet of Things
V2X	Vehicle-to-Everything
V2V	Vehicle-to-Vehicle
V2I	Vehicle-to-Infrastructure
V2P	Vehicle-to-Pedestrian
V2C	Vehicle-to-Cloud
Inter-V	Inter-Vehicle
OBU	On-Board Unit
RSU	Road-Side Unit
DoS	Denial of Service
DDoS	Distributed Denial of Service
KNN	K-Nearest Neighbor

SVM	Support Vector Machine
RF	Random Forest
DT	Decision Tree
ET	Extra Trees
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient-Boosting Machine
AdaBoost	Adaptive Boosting
LDA	Linear Discriminant Analysis
OCSVM	One-Class Support Vector Machine
FPR	False Positive Rate
FNR	False Negative Rate
ANN	Artificial Neural Network
MLP	Multi-Layer Perceptron
DNN	Deep Neural Network
DBN	Deep Belief Network
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
LSTM	Long Short-Term Memory
Bi-LSTM	Bidirectional LSTM
BiGAN	Bidirectional Generative Adversarial Network
ACGAN	Auxiliary Classifier GAN
AE	Autoencoder
CVAE	Conditional Variational Autoencoder
A-VAE	Attention-based Variational Autoencoder
Conv-AE	Convolutional Autoencoder
SA	Self-Attention
FL	Federated Learning
EC	Edge Computing
MDS	Misbehavior Detection System
LLM	Large Language Model
BSM	Basic Safety Message

References

1. Man, D.; Zeng, F.; Lv, J.; Xuan, S.; Yang, W.; Guizani, M. AI-Based Intrusion Detection for Intelligence Internet of Vehicles. *IEEE Consum. Electron. Mag.* **2021**, *12*, 109–116. [CrossRef]
2. Almehdhar, M.; Albaseer, A.; Khan, M.A.; Abdallah, M.; Menouar, H.; Al-Kuwari, S.; Al-Fuqaha, A. Deep Learning in the Fast Lane: A Survey on Advanced Intrusion Detection Systems for Intelligent Vehicle Networks. *IEEE Open J. Veh. Technol.* **2024**, *5*, 869–906. [CrossRef]
3. Rajapaksha, S.; Kalutarage, H.; Al-Kadri, M.O.; Petrovski, A.; Madzudzo, G.; Cheah, M. AI-Based Intrusion Detection Systems for In-Vehicle Networks: A Survey. *ACM Comput. Surv.* **2023**, *55*, 237. [CrossRef]
4. Karim, S.M.; Habbal, A.; Chaudhry, S.A.; Irshad, A. Architecture, Protocols, and Security in IoV: Taxonomy, Analysis, Challenges, and Solutions. *Secur. Commun. Netw.* **2022**, *2022*, 1131479. [CrossRef]
5. Liu, N. Internet of Vehicles: Your Next Connection. *WinWin Mag.* **2011**, *11*, 23–28.
6. Kaiwartya, O.; Abdullah, A.H.; Cao, Y.; Altameem, A.; Prasad, M.; Lin, C.T.; Liu, X. Internet of Vehicles: Motivation, Layered Architecture, Network Model, Challenges, and Future Aspects. *IEEE Access* **2016**, *4*, 5356–5373. [CrossRef]
7. Contreras-Castillo, J.; Zeadally, S.; Guerrero-Ibañez, J.A. Internet of Vehicles: Architecture, Protocols, and Security. *IEEE Internet Things J.* **2017**, *5*, 3701–3709. [CrossRef]
8. Liu, K.; Xu, X.; Chen, M.; Liu, B.; Wu, L.; Lee, V.C. A Hierarchical Architecture for the Future Internet of Vehicles. *IEEE Commun. Mag.* **2019**, *57*, 41–47. [CrossRef]
9. Ji, B.; Zhang, X.; Mumtaz, S.; Han, C.; Li, C.; Wen, H.; Wang, D. Survey on the Internet of Vehicles: Network Architectures and Applications. *IEEE Commun. Stand. Mag.* **2020**, *4*, 34–41. [CrossRef]
10. Gao, L.; Wu, C.; Du, Z.; Yoshinaga, T.; Zhong, L.; Liu, F.; Ji, Y. Toward Efficient Blockchain for the Internet of Vehicles with Hierarchical Blockchain Resource Scheduling. *Electronics* **2022**, *11*, 832. [CrossRef]

11. Wang, Y.; Ning, W.; Zhang, S.; Yu, H.; Cen, H.; Wang, S. Architecture and Key Terminal Technologies of 5G-Based Internet of Vehicles. *Comput. Electr. Eng.* **2021**, *95*, 107430. [CrossRef]
12. Samad, A.; Alam, S.; Mohammed, S.; Bhukhari, M. Internet of Vehicles (IoV) Requirements, Attacks and Countermeasures. In Proceedings of the 12th INDIACom—5th International Conference on Computing for Sustainable Global Development, New Delhi, India, 14–16 March 2018; pp. 1–4.
13. Bagga, P.; Das, A.K.; Wazid, M.; Rodrigues, J.J.P.C.; Park, Y. Authentication Protocols in Internet of Vehicles: Taxonomy, Analysis, and Challenges. *IEEE Access* **2020**, *8*, 54314–54344. [CrossRef]
14. Hasrouny, H.; Samhat, A.E.; Bassil, C.; Laouiti, A. VANET Security Challenges and Solutions: A Survey. *Veh. Commun.* **2017**, *7*, 7–20. [CrossRef]
15. Francillon, A.; Danev, B.; Capkun, S. Relay Attacks on Passive Keyless Entry and Start Systems in Modern Cars. In Proceedings of the 18th Network and Distributed System Security Symposium (NDSS), San Diego, CA, USA, 6–9 February 2011.
16. Bariah, L.; Shehada, D.; Salahat, E.; Yeun, C.Y. Recent Advances in VANET Security: A Survey. In Proceedings of the 2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), Boston, MA, USA, 6–9 September 2015; pp. 1–7.
17. Haddaji, A.; Ayed, S.; Chaari Fourati, L. IoV Security and Privacy Survey: Issues, Countermeasures, and Challenges. *J. Supercomput.* **2024**, *80*, 23018–23082. [CrossRef]
18. Sun, Y.; Wu, L.; Wu, S.; Li, S.; Zhang, T.; Zhang, L.; Cui, X. Attacks and Countermeasures in the Internet of Vehicles. *Ann. Telecommun.* **2017**, *72*, 283–295. [CrossRef]
19. Cao, H.; Xu, G.; He, Z.; Shi, S.; Xu, S.; Wu, C.; Ning, J. Unveiling the superiority of unsupervised learning on GPU cryptojacking detection: Practice on magnetic side channel-based mechanism. *IEEE Trans. Inf. Forensics Secur.* **2025**, *20*, 4874–4889. [CrossRef]
20. Cao, H.; Liu, D.; Jiang, H.; Luo, J. MagSign: Harnessing dynamic magnetism for user authentication on IoT devices. *IEEE Trans. Mob. Comput.* **2022**, *23*, 597–611. [CrossRef]
21. Ni, T.; Zhang, X.; Zuo, C.; Li, J.; Yan, Z.; Wang, W.; Xu, W.; Luo, X.; Zhao, Q. Uncovering user interactions on smartphones via contactless wireless charging side channels. In Proceedings of the 2023 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 22–24 May 2023; pp. 3399–3415. [CrossRef]
22. Ni, T.; Lan, G.; Wang, J.; Zhao, Q.; Xu, W. Eavesdropping mobile app activity via radio-frequency energy harvesting. In Proceedings of the 32nd USENIX Security Symposium (USENIX Security 23), Anaheim, CA, USA, 9–11 August 2023; pp. 3511–3528.
23. Ni, T.; Zhang, X.; Zhao, Q. Recovering fingerprints from in-display fingerprint sensors via electromagnetic side channel. In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security, New York, NY, USA, 26–30 November 2023; pp. 253–267.
24. Hu, P.; Zhuang, H.; Santhalingam, P.S.; Spolaor, R.; Pathak, P.; Zhang, G.; Cheng, X. Accear: Accelerometer acoustic eavesdropping with unconstrained vocabulary. In Proceedings of the 2022 IEEE Symposium on Security and Privacy (SP), Francisco, CA, USA, 23–25 May 2022; pp. 1757–1773.
25. Duan, D.; Sun, Z.; Ni, T.; Li, S.; Jia, X.; Xu, W.; Li, T. F2key: Dynamically converting your face into a private key based on COTS headphones for reliable voice interaction. In Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services, New York, NY, USA, 3–7 June 2024; pp. 127–140.
26. Zhou, Y.; Ni, T.; Lee, W.-B.; Zhao, Q. A survey on backdoor threats in large language models (LLMs): Attacks, defenses, and evaluation methods. *Trans. Artif. Intell.* **2025**, *3*, 3. [CrossRef]
27. Wang, J.; Ni, T.; Lee, W.-B.; Zhao, Q. A contemporary survey of large language model assisted program analysis. *arXiv* **2025**, arXiv:2502.18474. [CrossRef]
28. Alshammari, A.; Zohdy, M.A.; Debnath, D.; Corser, G. Classification Approach for Intrusion Detection in Vehicle Systems. *Wirel. Eng. Technol.* **2018**, *9*, 79–94. [CrossRef]
29. Yang, L.; Moubayed, A.; Hamieh, I.; Shami, A. Tree-Based Intelligent Intrusion Detection System in Internet of Vehicles. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.
30. Vuong, T.P.; Loukas, G.; Gan, D. Performance Evaluation of Cyber-Physical Intrusion Detection on a Robotic Vehicle. In Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, UK, 26–28 October 2015; pp. 2106–2113.
31. Albishi, O.A.; Abdullah, M. DDoS Attacks Detection in IoV Using ML-Based Models with an Enhanced Feature Selection Technique. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*, 814. [CrossRef]
32. Sumathi, S.; Rajesh, R.; Karthikeyan, N. DDoS Attack Detection Using Hybrid Machine Learning-Based IDS Models. *J. Sci. Ind. Res.* **2022**, *81*, 276–286. [CrossRef]
33. Halladay, J.; Cullen, D.; Briner, N.; Warren, J.; Fye, K.; Basnet, R.; Doleck, T. Detection and Characterization of DDoS Attacks Using Time-Based Features. *IEEE Access* **2022**, *10*, 49794–49807. [CrossRef]

34. Minawi, O.; Whelan, J.; Almehmadi, A.; El-Khatib, K. Machine Learning-Based Intrusion Detection System for Controller Area Networks. In Proceedings of the 10th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications, Miami, FL, USA, 16–20 November 2020; pp. 41–47.
35. Goncalves, F.; Macedo, J.; Santos, A. Intelligent Hierarchical Intrusion Detection System for VANETs. In Proceedings of the 2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Brno, Czech Republic, 25–27 October 2021; pp. 50–59.
36. Kosmanos, D.; Pappas, A.; Aparicio-Navarro, F.J.; Maglaras, L.; Janicke, H.; Boiten, E.; Argyriou, A. Intrusion Detection System for Platooning Connected Autonomous Vehicles. In Proceedings of the 2019 4th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Piraeus, Greece, 20–22 September 2019; pp. 1–9.
37. Sousa, B.; Magaia, N.; Silva, S. An Intelligent Intrusion Detection System for 5G-Enabled Internet of Vehicles. *Electronics* **2023**, *12*, 1757. [CrossRef]
38. Sharma, A.; Jaekel, A. Machine Learning Based Misbehaviour Detection in VANET Using Consecutive BSM Approach. *IEEE Open J. Veh. Technol.* **2021**, *3*, 1–14. [CrossRef]
39. Song, H.M.; Kim, H.R.; Kim, H.K. Intrusion Detection System Based on the Analysis of Time Intervals of CAN Messages for In-Vehicle Network. In Proceedings of the 2016 International Conference on Information Networking (ICOIN), Kota Kinabalu, Malaysia, 13–15 January 2016; pp. 63–68. [CrossRef]
40. Marchetti, M.; Stabili, D. Anomaly Detection of CAN Bus Messages through Analysis of ID Sequences. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Redondo Beach, CA, USA, 11–14 June 2017; pp. 1577–1583. [CrossRef]
41. Stabili, D.; Marchetti, M.; Colajanni, M. Detecting Attacks to Internal Vehicle Networks through Hamming Distance. In Proceedings of the 2017 AEIT International Annual Conference, Cagliari, Italy, 20–22 September 2017; pp. 1–6. [CrossRef]
42. Narayanan, S.N.; Mittal, S.; Joshi, A. Using Data Analytics to Detect Anomalous States in Vehicles. *arXiv* **2015**, arXiv:1512.08048. [CrossRef]
43. Avatefipour, O. Physical-Fingerprinting of Electronic Control Unit (ECU) Based on Machine Learning Algorithm for In-Vehicle Network Communication Protocol "CAN-BUS". Master's Thesis, University of Michigan-Dearborn, Dearborn, MI, USA, 2017. Available online: https://hdl.handle.net/2027.42/140731 (accessed on 14 April 2025).
44. Anzer, A.; Elhadef, M. A Multilayer Perceptron-Based Distributed Intrusion Detection System for Internet of Vehicles. In Proceedings of the 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC), Philadelphia, PA, USA, 18–20 October 2018; pp. 438–445.
45. Kang, M.J.; Kang, J.W. Intrusion Detection System Using Deep Neural Network for In-Vehicle Network Security. *PLoS ONE* **2016**, *11*, e0155781. [CrossRef] [PubMed]
46. Xun, Y.; Zhao, Y.; Liu, J. VehicleEIDS: A Novel External Intrusion Detection System Based on Vehicle Voltage Signals. *IEEE Internet Things J.* **2021**, *9*, 2124–2133. [CrossRef]
47. Zhang, L.; Shi, L.; Kaja, N.; Ma, D. A two-stage deep learning approach for CAN intrusion detection. In Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS), Novi, MI, USA, 7–9 August 2018; pp. 1–11.
48. Vitalkar, R.S.; Thorat, S.S. A Review on Intrusion Detection System in Vehicular Ad Hoc Network Using Deep Learning Method. *Int. J. Res. Appl. Sci. Eng. Technol.* **2020**, *8*, 1591–1595. [CrossRef]
49. Peng, R.; Li, W.; Yang, T.; Huafeng, K. An Internet of Vehicles Intrusion Detection System Based on a Convolutional Neural Network. In Proceedings of the 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Xiamen, China, 16–18 December 2019; pp. 1595–1599.
50. Desta, A.K.; Ohira, S.; Arai, I.; Fujikawa, K. Rec-CNN: In-Vehicle Networks Intrusion Detection Using Convolutional Neural Networks Trained on Recurrence Plots. *Veh. Commun.* **2022**, *35*, 100470. [CrossRef]
51. Nie, L.; Ning, Z.; Wang, X.; Hu, X.; Cheng, J.; Li, Y. Data-Driven Intrusion Detection for Intelligent Internet of Vehicles: A Deep Convolutional Neural Network-Based Method. *IEEE Trans. Netw. Sci. Eng.* **2020**, *7*, 2219–2230. [CrossRef]
52. Ashraf, J.; Bakhshi, A.D.; Moustafa, N.; Khurshid, H.; Javed, A.; Beheshti, A. Novel Deep Learning-Enabled LSTM Autoencoder Architecture for Discovering Anomalous Events from Intelligent Transportation Systems. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 4507–4518. [CrossRef]
53. Agrawal, K.; Alladi, T.; Agrawal, A.; Chamola, V.; Benslimane, A. NovelADS: A Novel Anomaly Detection System for Intra-Vehicular Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22596–22606. [CrossRef]
54. Khan, I.A.; Moustafa, N.; Pi, D.; Haider, W.; Li, B.; Jolfaei, A. An Enhanced Multi-Stage Deep Learning Framework for Detecting Malicious Activities from Autonomous Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 25469–25478. [CrossRef]
55. Zeng, Y.; Qiu, M.; Zhu, D.; Xue, Z.; Xiong, J.; Liu, M. DeepVCM: A Deep Learning Based Intrusion Detection Method in VANET. In Proceedings of the 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), High Performance and Smart Computing (HPSC), and Intelligent Data and Security (IDS), Washington, DC, USA, 27–29 May 2019; pp. 288–293.

56. Alladi, T.; Kohli, V.; Chamola, V.; Yu, F.R.; Guizani, M. Artificial Intelligence (AI)-Empowered Intrusion Detection Architecture for the Internet of Vehicles. *IEEE Wirel. Commun.* **2021**, *28*, 144–149. [CrossRef]
57. Wang, C.; Zhao, Z.; Gong, L.; Zhu, L.; Liu, Z.; Cheng, X. A distributed anomaly detection system for in-vehicle network using HTM. *IEEE Access* **2018**, *6*, 9091–9098. [CrossRef]
58. Loukas, G.; Vuong, T.; Heartfield, R.; Sakellari, G.; Yoon, Y.; Gan, D. Cloud-based cyber-physical intrusion detection for vehicles using deep learning. *IEEE Access* **2017**, *6*, 3491–3508. [CrossRef]
59. Banh, L.; Strobel, G. Generative Artificial Intelligence. *Electron. Mark.* **2023**, *33*, 63. [CrossRef]
60. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1–9.
61. Salehi, P.; Chalechale, A.; Taghizadeh, M. Generative Adversarial Networks (GANs): An Overview of Theoretical Model, Evaluation Metrics, and Recent Developments. *arXiv* **2020**, arXiv:2005.13178. [CrossRef]
62. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.U.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
63. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2013**, arXiv:1312.6114.
64. Asperti, A.; Evangelista, D.; Loli Piccolomini, E. A Survey on Variational Autoencoders from a Green AI Perspective. *SN Comput. Sci.* **2021**, *2*, 301. [CrossRef]
65. Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; Ganguli, S. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics. In Proceedings of the International Conference on Machine Learning (ICML), Lille, France, 7–9 July 2015; pp. 2256–2265.
66. Xie, Y.; Yu, J.; Ranneby, B. A General Autoregressive Model with Markov Switching: Estimation and Consistency. *Math. Methods Stat.* **2008**, *17*, 228–240. [CrossRef]
67. Khalil, A.; Farman, H.; Nasralla, M.M.; Jan, B.; Ahmad, J. Artificial Intelligence-Based Intrusion Detection System for V2V Communication in Vehicular Adhoc Networks. *Ain Shams Eng. J.* **2024**, *15*, 102616. [CrossRef]
68. Poongodi, M.; Hamdi, M. Intrusion Detection System Using Distributed Multilevel Discriminator in GAN for IoT System. *Trans. Emerg. Telecommun. Technol.* **2023**, *34*, e4815. [CrossRef]
69. Shu, J.; Zhou, L.; Zhang, W.; Du, X.; Guizani, M. Collaborative Intrusion Detection for VANETs: A Deep Learning-Based Distributed SDN Approach. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 4519–4530. [CrossRef]
70. Chen, X.; Xiao, K.; Luo, L.; Li, Y.; Chen, L. GAN-IVDS: An Intrusion Detection System for Intelligent Connected Vehicles Based on Generative Adversarial Networks. In Proceedings of the 2023 8th International Conference on Data Science in Cyberspace (DSC), Hefei, China, 18–20 August 2023; pp. 237–244.
71. Qiu, Y.; Misu, T.; Busso, C. Unsupervised Scalable Multimodal Driving Anomaly Detection. *IEEE Trans. Intell. Veh.* **2022**, *8*, 3154–3165. [CrossRef]
72. Seo, E.; Song, H.M.; Kim, H.K. GIDS: GAN Based Intrusion Detection System for In-Vehicle Network. In Proceedings of the 2018 16th Annual Conference on Privacy, Security and Trust (PST), Belfast, Ireland, 28–30 August 2018; pp. 1–6.
73. Wang, X.; Xu, Y.; Xu, Y.; Wang, Z.; Wu, Y. Intrusion Detection System for In-Vehicle CAN-FD Bus ID Based on GAN Model. *IEEE Access* **2024**, *12*, 82402–82412. [CrossRef]
74. Xie, G.; Yang, L.T.; Yang, Y.; Luo, H.; Li, R.; Alazab, M. Threat Analysis for Automotive CAN Networks: A GAN Model-Based Intrusion Detection Technique. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 4467–4477. [CrossRef]
75. Zhao, Q.; Chen, M.; Gu, Z.; Luan, S.; Zeng, H.; Chakrabory, S. CAN Bus Intrusion Detection Based on Auxiliary Classifier GAN and Out-of-Distribution Detection. *ACM Trans. Embed. Comput. Syst.* **2022**, *21*, 45. [CrossRef]
76. Nguyen, T.P.; Nam, H.; Kim, D. Transformer-Based Attention Network for In-Vehicle Intrusion Detection. *IEEE Access* **2023**, *11*, 55389–55403. [CrossRef]
77. Li, M.; Han, D.; Li, D.; Liu, H.; Chang, C.C. MFVT: An Anomaly Traffic Detection Method Merging Feature Fusion Network and Vision Transformer Architecture. *EURASIP J. Wirel. Commun. Netw.* **2022**, *2022*, 39. [CrossRef]
78. Cobilean, V.; Mavikumbure, H.S.; Wickramasinghe, C.S.; Varghese, B.J.; Pennington, T.; Manic, M. Anomaly Detection for In-Vehicle Communication Using Transformers. In Proceedings of the IECON 2023—49th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 16–19 October 2023; pp. 1–6.
79. Lai, Q.; Xiong, C.; Chen, J.; Wang, W.; Chen, J.; Gadekallu, T.R.; Hu, X. Improved Transformer-Based Privacy-Preserving Architecture for Intrusion Detection in Secure V2X Communications. *IEEE Trans. Consum. Electron.* **2023**, *70*, 1810–1820. [CrossRef]
80. Liu, Z.; Xu, H.; Kuang, Y.; Li, F. SVMDFormer: A Semi-Supervised Vehicular Misbehavior Detection Framework Based on Transformer in IoV. In Proceedings of the 2023 IEEE 43rd International Conference on Distributed Computing Systems (ICDCS), Hong Kong, China, 18–21 July 2023; pp. 887–897.
81. Hamhoum, W.; Cherkaoui, S. MistralBSM: Leveraging Mistral-7B for Vehicular Networks Misbehavior Detection. *arXiv* **2024**, arXiv:2407.18462. [CrossRef]

82. Li, X.; Fu, H. SecureBERT and LLAMA 2 Empowered Control Area Network Intrusion Detection and Classification. *arXiv* **2023**, arXiv:2311.12074. [CrossRef]
83. Nwafor, E.; Olufowobi, H. CanBERT: A Language-Based Intrusion Detection Model for In-Vehicle Networks. In Proceedings of the 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), Nassau, Bahamas, 12–14 December 2022; pp. 294–299.
84. Aldossary, S.M. Smart Vehicles Networks: BERT Self-Attention Mechanisms for Cyber-Physical System Security. *Int. J. Syst. Assur. Eng. Manag.* **2023**, 1–9. [CrossRef]
85. Tang, W.; Li, D.; Fan, W.; Liu, T.; Chen, M.; Dib, O. An Intrusion Detection System Empowered by Deep Learning Algorithms. In Proceedings of the 2023 IEEE International Conference on Dependable, Autonomic and Secure Computing, Pervasive Intelligence and Computing, Cloud and Big Data Computing, Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Abu Dhabi, United Arab Emirates, 14–17 November 2023; pp. 1137–1142.
86. Alkhatib, N.; Mushtaq, M.; Ghauch, H.; Danger, J.L. CAN-BERT Do It? Controller Area Network Intrusion Detection System Based on BERT Language Model. In Proceedings of the 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, 5–8 December 2022; pp. 1–8.
87. Monshizadeh, M.; Khatri, V.; Gamdou, M.; Kantola, R.; Yan, Z. Improving Data Generalization with Variational Autoencoders for Network Traffic Anomaly Detection. *IEEE Access* **2021**, *9*, 56893–56907. [CrossRef]
88. Aslam, N.; Kolekar, M.H. A-VAE: Attention-Based Variational Autoencoder for Traffic Video Anomaly Detection. In Proceedings of the 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 7–9 April 2023; pp. 1–7.
89. Wei, Y.; Jang-Jaccard, J.; Sabrina, F.; Singh, A.; Xu, W.; Camtepe, S. AE-MLP: A Hybrid Deep Learning Approach for DDoS Detection and Classification. *IEEE Access* **2021**, *9*, 146810–146821. [CrossRef]
90. Kim, T.; Kim, J.; You, I. An Anomaly Detection Method Based on Multiple LSTM-Autoencoder Models for In-Vehicle Network. *Electronics* **2023**, *12*, 3543. [CrossRef]
91. Hoang, T.N.; Kim, D. Detecting In-Vehicle Intrusion via Semi-Supervised Learning-Based Convolutional Adversarial Autoencoders. *Veh. Commun.* **2022**, *38*, 100520. [CrossRef]
92. Wei, P.; Wang, B.; Dai, X.; Li, L.; He, F. A Novel Intrusion Detection Model for the CAN Bus Packet of In-Vehicle Network Based on Attention Mechanism and Autoencoder. *Digit. Commun. Netw.* **2023**, *9*, 14–21. [CrossRef]
93. *ISO/SAE 21434:2021*; Road vehicles—Cybersecurity engineering. International Organization for Standardization: Geneva, Switzerland, 2020.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.