Review Article: Problems and the Approaches of Machine Learning in Vehicle Ad Hoc Networks

Skala Hassan Hussen1,* and Marwan Aziz Mohammed2

1 College of Engineering, Department of Software and Informatics Engineering, Salahaddin University, Erbil, Iraq
2 College of Engineering, Department of Computer engineering, Knowledge university, Erbil, Iraq

Abstract. In recent years, there has been a notable surge in research interest in vehicular ad-hoc networks (VANETs) due to advancements in wireless communication technology and the vehicle sector. Vehicles to vehicles (V2V) and vehicles to infrastructure comprise a vehicular network. The potential machine learning (ML) method can offer practical solutions for various application fields. Machine learning is a technique where a system uses data that has already been processed to learn from and improve itself automatically. Vehicular networks are a significant application domain where ML-based techniques are highly helpful in solving various issues. Vehicular nodes and infrastructure communicating wirelessly are susceptible to many kinds of assaults. Intelligent transportation systems (ITS) rely heavily on vehicle ad hoc networks (VANETs). These methods enable effective supervised and unsupervised learning of the acquired data, hence accomplishing the goal of VANETs. Because of identifying security concerns in-vehicle networks from source to destination, this evaluation attempts to apply it. We outlined the problems with traffic, safety, and communication in VANET systems, discussed whether or not they could be implemented, and investigated the potential solutions provided by machine learning techniques.

1. Introduction

The original equipment Automobile manufacturers (OEMs) are making new models because people want more ease and safety in their cars. There are now self-driving cars on the market that can connect to a VANET transmission network (Marwah and Jain, 2022). IEEE 802.11, also called VANET transmission, is used for the wireless connection between the cars and the Roadside Unit (RSU). A study by Deep Singh, Rawat, and Bonnin (2014) says that VANET can communicate in two ways: vehicle-
to-vehicle (V2V) and vehicle-to-infrastructure (V2I) [1]. in addition, With the rise of smart cities, intelligent transportation is becoming more and more important. Traffic flow tracking and congestion control are two of the most important parts of a smart city. The auto industry, governments, and universities all think that the idea of direct vehicle-to-vehicle kinematic data exchange over an ad hoc network environment called the vehicle ad hoc network (VANET) is a good one for the future of intelligent transportation systems (ITS), which will make our almost-congested highways safer and more efficient. There is a greater need to study VANET, which includes cars with built-in devices like GPS [2]. More and more people are interested in VANET study. To make V2V and V2I communication easier, cars are being equipped with Onboard Units (OBU) that have sensors, other wireless devices, and the Global Positioning System (GPS) [3]. Because cars can move around in the vehicular ad-hoc network, there are times when communication breaks down, the network gets crowded, and links between VANET parts become unstable. In order for VANETs to communicate, authentication systems need to be able to grow as needed, be stable, and not cost a lot to run. Software Defined Networking (SDN) based on 5G networks is now part of VANET. This means that efficient routing methods are needed [4]. Original Equipment Automobile manufacturers, or OEMs, are developing new models in response to consumers' growing car safety and comfort demands. Autonomous vehicles can currently communicate over a VANET communication network. VANET communication, defined by IEEE 802.11, is the foundation for the wireless communication between cars and the Roadside Unit (RSU) [5]. Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) are the two possible modes of communication in VANET. As well as smart cities take shape, intelligent mobility is becoming more and more critical. Controlling traffic congestion and monitoring traffic flow are two of a smart city's main characteristics. The potential for direct vehicle-to-vehicle kinematic data exchange over the vehicle ad hoc network (VANET), an ad hoc network environment, is seen by governments, the auto industry, and academia as a promising concept for the realization of intelligent transportation systems (ITS) in the future, ultimately achieving efficiency and safety in our nearly congested motorways. Research is becoming increasingly necessary for VANET, which comprises automobiles with onboard electronics like GPS [6].Growing in popularity is VANET research, which uses vehicles outfitted with Onboard Units (OBU) with sensors, additional wireless devices, and the Global Positioning System (GPS) to enable V2V and V2I communication. VANET components have unreliable connections, network congestion, and communication breakdowns due to car mobility within the vehicular ad hoc network. Authentication systems must be scalable, dependable, and have a low processing cost to facilitate communication over VANETs. Effective routing protocols are needed now that VANET uses Software Defined Networking (SDN) based on 5G networks [7].

1.1 VANET's machine learning models and approaches

Forwarding a packet from one vehicle to another is called routing. The network topology or the locations of the cars within the VANET may generally be used to determine the forwarding. In general, location-based routing protocols use the vehicles' location and direction to move packets from one vehicle to another. For VANET communication, various protocols are available. When developing routing
protocols, most consider fundamental criteria such as throughput, average transmission latency, packet delivery ratio, number of hops needed for transmission, etc. VANETs can benefit from the communication tools that machine learning offers. Through the use of machine learning algorithms, the creation, exchange, and transmission of data within VANETs can be helpful. Supervised Learning (SL), Unsupervised Learning (UL), and Reinforcement Learning (RL) are the three categories into which machine learning approaches fall [8].

![Machine Learning in VANET](image)

**Fig. 1.** Machine Learning in VANET

Several machine learning techniques, such as k-nearest neighbors (KNN), Support Vector Machine (SVM), k-means clustering, Naive Bayes (NB), Convolutional Neural Networks (CNN), and Artificial Neural Networks (ANN), can be used to look at the data flow in the VANET system and find problems like these. In addition, safety is one of the main goals of the smart transportation system. Finding items and roadblocks is the most important part of safety provisioning. The k-NN and CNN algorithms work better when using visual data to find obstacles [9]. When it comes to Gbps lines, which are needed for 5G-VANET data sharing, millimeter bands work best. Algorithms in machines can choose the right beam for data flow.

Modern automobiles are becoming more and more popular. These cars offer more advanced features including multimedia systems, integrated wireless access systems, and environmental awareness in addition to navigation and global positioning systems (GPS), which lower the risk of auto accidents and improve user experience [10]. Moreover, there is a lot of interest in improving the efficiency of vehicle communications. ITS aims to accomplish this by improving the quality, safety, and efficiency of transportation infrastructure and vehicles through the use of information and communication technology (ICT). In order to lessen traffic problems, ITS also gives clients the most recent information on route specifics and traffic congestion.

As a result, it uses less fuel and releases fewer pollutants into the atmosphere. Many countries are considering the applications of ITS. Additionally, the ITS contributes to the efficient use of the infrastructure and road safety. Mobile notifications, traffic control, vehicle safety, maintenance and construction management, and vehicle
detection are among the functions covered by Intelligent Transportation Systems (ITS).

ITS lowers costs and safeguards both the driver's and the passenger's health by coordinating among the cars and providing timely alerts. Information and communication technologies (ICT) objective and intelligent transportation systems (ITS) is to provide sustainable and cost-effective transportation by creating cutting-edge applications and services that optimize travel times and energy usage. All forms of communication within cars, between vehicles, and between vehicles and roadside infrastructure are supported by ITS in a variety of communication scenarios. There are two categories for vehicular communications: V2V and V2I [11]. Cellular network connection and vehicle-to-roadside (V2R) communication are also included in V2I. An example of a vehicle ad hoc network (VANET) is vehicle-to-vehicle (V2V) communications, created between automobiles to exchange data, such as safety information. Information is shared between a vehicle's onboard unit (OBU) and roadside unit (RSU) via vehicle-to-vehicle radio (V2R). Information is shared in V2I between the OBU and the RSU or a cellular network [12].

Fig. 2 Vehicle-to-Vehicle

1.2 Machine Learning Model and Approaches in VANET

Machine learning is used by Vehicular Ad-hoc Networks (VANETs) to enhance safety, traffic control, and vehicle-to-roadside equipment communication efficiency. Numerous machine learning techniques, which are enumerated above, can be used with VANETs:

1.3 Supervised Education:

Classification and Regression: This method forecasts traffic flow, vehicle behavior, and accident risks using models such as logistic regression and support vector machines. Trees of Decisions: used to choose routes depending on factors such as time, distance, and speed.

1.3.1 Unmonitored Education

Clustering: Methods such as K-means or DBSCAN identify clusters of vehicles for efficient information exchange or identify anomalous movement patterns that may indicate threats. Dimensionality reduction is the process of reducing the number of
variables using methods such as Principal Component Analysis, which facilitates easier data handling.

1.4 Learning through Reinforcement:

Algorithms like DQN and Q-Learning enable real-time decision-making with optimal routing and adaptive traffic control. Multiagent systems: automobiles can improve their driving techniques to reduce traffic by learning from their surroundings.

1.5 Artificial Intelligence:

Convolutional Neural Networks: Quite useful for image-based applications, such as lane detection, traffic sign identification, and vehicle identification. Both Long Short-Term Memory Networks (LSTMs) and Recurrent Neural Networks (RNNs) can be used to estimate traffic conditions based on past data. Federated Education by enabling cars to train a joint model without ever disclosing their data, V2X preserves privacy and reduces the need for data transfer. Edge computing processes this kind of data close to real-time at the edge, minimizing reliance on central servers; this is helpful, for instance, in preventing collisions. Although there are some significant variations among the models, these methods essentially use the data generated by the infrastructure and the vehicles to improve the VANET's efficiency, safety, and dependability. The model that is used depends on the specific application, the availability of data, and the available computing power.

2. Vehicular communications

Modern automobiles are becoming more and more popular. In addition, in the global positioning system (GPS), there are more advanced features in cars like environmental awareness, integrated wireless access systems, and multimedia systems, which decrease the risk of conflict between vehicles and user experience. Additionally, communication between cars increases effectiveness. By using the technology in communication and information. To enhance transportation structure and vehicles, dependency safety, and efficiency. Additionally, to provide practical ITS aims and transportation by utilizing application services that decrease energy usage and travel times. All connectivity with vehicles and between cars and roadside infrastructure is provided by ITS in different communication situations. The vehicular communication in classification is shown in Figure 2. Vehicle communications are in two categories, which are V2V and V2I. The connection between cellular connection networks and cars to roadside communication is also called V2I. An example of a vehicle ad hoc network (VANET) is vehicle-to-vehicle (V2V) communications, created between automobiles to exchange data, such as safety information. Information is shared between a vehicle's onboard unit (OBU) and roadside unit (RSU) via vehicle-to-vehicle radio (V2R). Information is shared in V2I between OBU, the RSU, or a cellular network (Amalia et al., 2023).
3. Machine Learning Aspect of Routing VANET

VANETs use both position-based and topology-based routing protocols. Position-based protocols are grouped into V2I and V2V routing protocols, while topology-based protocols are grouped into proactive and reactive protocols. As the name suggests, topology-based routing decides how to route traffic based on the way that networks talk to each other [13]. This means that every node in the network has access to node topological information, even if they are not talking to each other. This is called a "proactive approach." It's also known as "table-driven." Reactive protocols, on the other hand, only show routing information to a destination when contact is needed. Trouble with reactive tactics is that they make it take longer to figure out how to get from one place to another [14].

![Diagram of VANET Routing Protocol]

**Fig. 3** VANET Routing Protocol

4. Security Attacks and Requirements

The needs and dangers of vehicle network security are covered in this section. It offered a taxonomy of security threats at various points in the vehicular network. As a result, they divide the assaults into four (4) categories: infrastructure-based, sensor-based, hardware or software (HW/SW)-based, and wireless communication-based in Figure 3 [15].
Fig. 4 Security Attack in Vehicular Network

4.1 Hardware/Software-based Attacks

The hardware and software systems of the VANET network are vulnerable to various attacks that aim to breach specific security criteria. The following lists several frequent attacks[16].

4.2 Timing Attack

This attack is side-channeled. It attempts to undermine a system's cryptographic algorithm. Fig. 4. Taxonomy of automotive network assaults by examining the temporal data needed to carry out the attack.

4.2.1 Infrastructure-based Attacks

The attacks that compromise the system at the infrastructure level are explained in this section. Repudiation Attack: This attack occurs at the application layer when malicious manipulations cause a system to lose control over the tracking of nodes and action logs.
4.2.2 Routing Attack

The routing procedure malfunctions as a result of this attack. The many types of routing assaults are further categorized based on the various levels at which the routing process malfunctions within a network. Hardware-wise, a network's routing tables may be impacted by a node's presence and location not being communicated [17].

4.2.3 GPS Spoofing

To track a global positioning system (GPS), this attack creates fictitious signals surrounding the car, which the GPS sensor picks up and stores as fictitious coordinates, confusing the vehicular networks' regular operation.

4.2.4 Spamming Attack

According to the attacker uses this approach to transmit many spam messages, which uses up network resources and causes data transmission delays.

5. Secure Communications

The blue rectangle that spans all of the layers in Figure 3 represents the security implementation in this design. It offers a variety of communication modalities so that organizations can choose the one that best fits their needs. For instance, one of the VPKIbrID modes employs PKI, better suited for unicast communications, in which a message is conveyed to a single recipient. The other employs ABE, which is helpful when a message is intended for several recipients (Slama, Alaya and Zidi, 2022) The sender entity may encrypt the message for many entities using VPKIbrID-ABE by using the corresponding attributes of each entity. Given that most communications in this scenario will have many targets, at least when between clusters, the VPKIbrID-ABE appears to be the most appropriate solution. But compared to the PKI mode, this mode is heavier and slower. Nevertheless, key cashing is possible with the VPKIbrID, greatly enhancing its functionality. The literature review on routing strategies and associated remarks reveals diverse approaches and their respective strengths and weaknesses. Propose a consistent clustering algorithm for classifying suspicious nodes, although it lacks consideration for crucial metrics like energy levels, potentially impacting overall network performance. introduce a Particle Swarm Optimization (PSO) based algorithm, albeit criticized for its time-intensive nature in finding optimal costs. Slama, Alaya, and Zidi (2022) support D&PMV, pointing out that even with its higher processing time and end-to-end latency, it is effective in identifying and stopping malicious nodes. In their exploration of deep network techniques for DDoS attack detection, Weber, Neves, and Ferreto (2021) observe that the use of additional fields in control packets for cryptographic algorithms causes routing cost and delay escalation. suggests a Support Vector Machine (SVM) method that uses a variety of machine learning methods, but runs into similar problems with lengthy repetitions. All of these studies
highlight how crucial it is to balance algorithmic complexity with real-world factors like processing time and end-to-end delay when designing routing techniques.

6. The Basic Concept of The Proposed Secure Routing and Clustering Protocol

In this subsection, the study outlines the basic concept of our recommended routing protocol. To stop a blackhole attack, the recommended routing protocol can be divided into two parts: clustering using the DLC protocol and routing using the DLSR protocol. These procedures are summed up as follows:

6.1 Routing Process

In order to find the target node. A blackhole node can send a fake RREP to make the source node think it has the best route. In order to get around this, our suggested DLSR strategy uses DL to find and pick the best way to avoid blackhole attacks. When a node gets an RREP packet, it sends the relevant feature data to the DNN model. The model takes this information into account and then sends an output that tells us whether the data came from a regular node or a blackhole node. Based on the fitness number, the routing protocol can then choose on the fly whether to run in secure routing mode or regular routing mode. Here, too, we use DL to find the best weight for the exercise function.

6.2 Clustering Process

In the suggested DLC protocol, network nodes are put into clusters based on data they share, such as their distance, speed, direction, and how much energy they still have. Which CH to use is based on which has the largest amount of leftover energy. DL is used to make the fitness function's weights work better so that the best CH can be chosen. This lets a node join a cluster as CM and send data smoothly. The DNN model uses inputs like cosine distance, cosine similarity, and leftover energy to figure out the best weight for the fitness function. It is also thought to improve route communication and lower the amount of control that needs to be done during clustering.

7. Methodology of VANET simulator

VANET simulation tools Network and mobility simulations are used to create VANET simulators. While mobility simulators handle the movement of individual nodes or their mobility, network simulators handle the modeling of communication protocols and message exchange. This section outlines the primary VANET simulators documented in the literature, emphasizing their features and architecture. Our search used widely used academic databases and search engines, such as IEEE Explorer, ACM Digital Library, Science Direct, and Google Scholar. Papers proposing a simulator or conducting a comparative analysis of VANET simulators
were considered. Furthermore, we meticulously examined their references. During this procedure, we reviewed about 20 papers in total. We also employed the Google Search Engine to locate proprietary VANET simulations that aren't always utilized by academics. Our selection of simulators is compiled in Table 2. We were able to locate some proprietary simulators, even though the majority are open-source. Our investigation is limited to simulators released after 2015 (shown in Table 2 with a gray backdrop). One of our analysis criteria in this work is that obsolete tools are likely to be discontinued and, as a result, are unlikely to assist the most recent developments in VANET research. It is cited for a thorough examination of earlier simulators.

Simulation tools play a crucial role in understanding and analyzing complex traffic scenarios, aiding in the development and validation of transportation systems. Among the notable tools available, SUMO stands out for its capability to handle real-life road networks, employing Python for programming and supporting both microscopic and mesoscopic simulation paradigms with QSim and JDEQSIm. MATSIM, utilizing Visual Basic, offers a time-continuous model suitable for highway and freeway systems, while PTV Vissim, also built on Visual Basic, focuses on microscopic simulation with a time-discrete car following model, operating on a demand and supply concept. AIMSUN, requiring no scripting, specializes in microscopic simulation of Connected and Autonomous Vehicles - Mobility as a Service, providing realistic representations of vehicle interactions. Lastly, CORSIM, utilizing C++ and Python, is tailored for large-scale design and validation of path-planning algorithms for self-driving vehicles, contributing to the advancement of Mobility as a Service and Demand Responsive Transportation.

7.1 Challenges in VANET

Security concerns: Machine learning is helpful in the following areas: defending against cyberattacks that compromise the security of VANETs. The ML-based technique has security flaws since it may produce unexpected results. Thus, even if ML has made remarkable progress in the VANET sectors, significant advancements in the security and robustness of machine learning-based techniques are still needed.

7.2 High Energy and Computing Requirements

Applications based on machine learning require a lot of computing power to run intricate and detailed models and processes. To operate and accomplish the necessary goals, WSNs and VANETs with high mobility nodes and traffic must consume significant energy. In addition to ML, WSNs and VANETs require lightweight and energy-efficient models.

7.3 Localization

While real-time applications require three-dimensional space, current localization techniques only handle two-dimensional space. Improving localization algorithms for three-dimensional situations is essential for VANETs, mobile and static.
7.4 High energy and Computing Requirements

Machine learning applications need a lot of processing power to operate complex and comprehensive models and processes. High mobility node and traffic WSNs and VANETs need to use a lot of energy to function and achieve the required objectives. WSNs and VANETs demand lightweight and energy-efficient models in addition to machine learning.

Different protocols address different use cases in the realm of routing protocols in research applications; each protocol has a routing mechanism and accompanying limitations. GPRS is a unicast system that is used for performance evaluation in urban environments, although its packet delivery ratio (PDR) is low. OLDER, which is used to compare different VANET protocols, uses geo casting but has higher end-to-end latency because of varying traffic density and topology. Targeting dependable packet routing in city environments, VADD uses unicast but has higher end-to-end latency. DSR employs unicast as well, but it has a low PDR in order to provide end-to-end Quality of Service (QoS) and transmission reliability.

Through broadcasting, A-STAR enables timely communication across large distances, but as traffic density changes, so do the network delays. With a low PDR, BROADCOM, which prioritises transmission dependability and end-to-end QoS, confronts difficulties. Inconsistent routing paths cause greater delays for ROVER, which guarantees packet routing with QoS for VANET. DV-CAST, intended for the transmission of emergency messages, is restricted to highway networks by means of geocasting. AODV uses unicast packet transmission and is devoted to inter-vehicle communication in dynamic VANET systems.

DOLPHIN is a broadcasting technique used to assess performance in urban environments assuming the presence of vehicles. DRG uses geo casting to enable timely communication over wide areas, but it ignores fluctuations in data traffic. MDDV uses broadcasting to promote group cooperative driving in highway scenarios, however its PDR is low.

GPRS, specifically for emergency message dissemination on highways, utilizes geo casting but also faces a low PDR. Lastly, PMB, efficient for reliable data dissemination, employs unicast but is deemed unsuitable for time-critical safety packet transmission in dynamic VANET environments.

Furthermore, security is a thoroughly studied problem with various solutions, including situation-modeling and ID-based techniques, in addition to the widely used and conventional Public Key Infrastructure (PKI). On the other hand, using IDSs, specifically intelligent IDSs, is a more modern technique that has gained increased interest since 2010.

The utilization of different network simulators (Net Sim) in tandem with various attack scenarios presents a multifaceted landscape in cybersecurity research. In the context of DoS attacks, NS2 employs neural networks for anomaly detection, strategically placing its focus on access points. In the meantime, MATLAB works with VANET Mobissimo in the domain of vehicular ad hoc networks (VANETs) and packet dropping assaults, employing Support Vector Machines (SVM) for detection and concentrating on watchdog mechanisms inside automobiles. While NS3 uses SUMO as its traffic simulator and Naive Bayes and Logistic Regression methods for anomaly detection, it focuses only on individual cells and cars. In contrast, NS3
handles a wider range of threats, such as DoS, R2L, and U2R. Furthermore, by utilizing SUMO and SVM for detection in vehicular environments, MATLAB and Net Sim work together to combat more complex attacks like wormhole, selective forwarding, and packet drop, emphasizing the significance of addressing malicious packet behaviors and learning automata in base stations. This thorough approach emphasizes how important it is to combine different simulators, machine learning methods, and detection methodologies in order to successfully counteract changing security.

A comparison of the most recent VANET surveys demonstrates a range of approaches and topics covered in multiple publications. Examine a range of variables, highlighting their significance in improving safety, vehicle monitoring, content delivery, and offloading in VANET contexts. These variables include road visibility conditions, walking distance, parking time, vehicle speed, and queue length. Marwah and Jain (2022) explore single-hop and multi-hop broadcast routing techniques, which are crucial for traffic control, safety applications, and resolving routing issues in VANETs. Focus on the volume, speed, flow, occupancy, and weather of the traffic. The use of machine learning (ML) in a Cognitive Ratio (CR) based VANET is discussed by Weber, Neves, and Ferreto (2021), with a focus on resource allocation, traffic congestion, security concerns, and accidents on the road. Use the NGSIM traffic dataset to identify communication misconduct, which is necessary to maintain VANET security. Emphasis on speed and video monitoring, which help with vehicle identification, tracking, and classification and are essential for maintaining efficiency and safety in VANET operations. Every post offers a different viewpoint, which together improve our knowledge and progress VANET applications and technologies.

8. Conclusions

In this paper, they looked into using machine learning to increase VANET efficiency, paying special emphasis to communication, safety, and traffic. We analyse the convergence of machine learning and VANET and track their evolution in vehicle networks over time. We also examine additional VANET applications that can leverage machine learning concepts, including data control, traffic monitoring, driver assistance, misbehavior detection, routing, and collision alarms. According to our research, SVM is frequently utilized for classification in applications ranging from misbehavior detection to collision detection, whereas neural network techniques are commonly employed in traffic management. Furthermore, deep learning and k-NN work well to help cars avoid collisions. These findings show how machine learning has much potential in VANET, especially in various AI-related fields. This study recommended utilizing machine learning algorithms to detect other network problems or issues, which helps collect multiple data from vehicles.
References


