Attacks on the Vehicle Ad-hoc Network from Cyberspace

Anas Alwasel[1], Shailendra Mishra[2], Mohammed AlShehri[3]
Department of Information Technology-College of Computer and Information Sciences,
Majmaah University, Majmaah, 11952, Saudi Arabia[1, 3]
Department of Computer Engineering-College of Computer and Information Sciences,
Majmaah University, Majmaah, 11952, Saudi Arabia[2]

***Abstract*—The emergence of Vehicle Ad hoc Networks (VANET) in 2003 has brought about a significant advancement in mobile phone networks and VANETs enable cars on the road to communicate with each other and the infrastructure on the street through a set of sensors and Intelligent Transport Systems (ITS). However VANETs are a low-level trust environment, making them vulnerable to misbehavior attacks and abnormal use. Thus, it is crucial to ensure that VANET systems and applications are secure and protected from cyber-attacks. This research aims to identify security challenges and vulnerabilities in VANET and proposes an algorithm that checks vehicle identity, location, and speed to detect and classify suspicious behavior. The research involves a study of the structures, architecture, and applications using VANET technology, the interconnection processes between them, and the types of architecture, layers, and applications that can pose a high risk. The research also focuses on the Confidentiality, Integrity and Availability (CIA) information security triangle and develops a program that uses machine learning to classify and analyze risks, attacks. The proposed algorithm provides security and safety for everyone on the road by identifying harmful behaviors of vehicles through knowledge of their location and identity. Overall, this research contributes to the development of a stable and secure Vehicular ad hoc network environment, enabling the integration of VANET security with smart cities.**

Keywords*—Vehicular Ad hoc Network (VANET); Mobile Ad hoc Network (MANET); machine learning; random forest; linear regression

I. INTRODUCTION

The Internet of Things, machine learning, and artificial intelligence have gained immense importance. Vehicular ad hoc network (VANET) is being used in various sectors, including business, industry, and military, and it will soon be available in the civil sector as well. Since machines will replace humans in performing tasks, it is crucial to ensure that they can perform their duties swiftly and accurately to avoid errors.

The underlying chapter is based on providing an overview of different areas that form the basis of this research. The areas that will be covered in this chapter include providing background of the research along with performing an evaluation of the problem that is being studied. Apart from that, the chapter will provide illustrations about the key definitions and the questions of research that are intended to be studied through the completion of this study.

The topic that is under consideration is about identifying Vehicle Ad hoc Networks and how to make this environment safer and less risky. The purpose of choosing this particular topic is because of the importance of the telecom sector and the growth within the level of technology that has been witnessed over the years in this particular industry. However, there have been concerns related to the safety of the use of this technology as well which has created the need to study this particular problem in a detailed manner.

Network and communication scientists are also dedicated to developing these technologies in a way that facilitates the process of their use and reduces errors in them with the ability to transfer data to many networks at high speed. One of the areas that use technology is the Intelligent Transport System (ITS). As per WHO, approximately 1.35 million people are killed each year worldwide by road crash accidents, costing countries 3% of the Gross Domestic Product (GDP) [1]. In Saudi Arabia, the 2019 Saudi Ministry of Interior statistical report [1] reported a total number of 352,646 road crash accidents, with a mean of 40.2 road crash accidents each hour. Of all road accidents, the number of people injured in the accidents was 17,295, and the number of people injured was 30,217. On the other hand, the number of car accidents where people died was 4,780, and the number of actual deaths was 6,025. The report showed that more car accidents (60.91 %) occurred within cities than outside cities. Most (61.08 %) occur in the morning compared to the evening. These statistics highlight the importance of activating ITS applications to support drivers to make the right decisions to decrease road accidents. And we can say that these vehicles can communicate with each other on the road through a connected network of smart systems and applications, and this is what is done with VANET technology, as this technology focuses on road safety and reducing accidents with more efficiency. This technology is classified as a type of wireless network that falls under the umbrella of Mobile Ad hoc Networks (MANETs)[2].

Privacy is also a significant concern in Vehicular ad hoc network (VANET)s, as they can potentially track the movements of vehicles [31]. VANETs must also scale up quickly and efficiently to support large numbers of vehicles on the road. Additionally, interoperability can be challenging due to different communication protocols used by different vehicles [32].

A. Vehicular Ad Hoc Network (VANET) Architecture

Vehicular ad hoc network (VANET)s rely on a reliable infrastructure, which can be costly and challenging to maintain in some areas. Fig. 2 shows the components of VANET architecture in this technology [4], which are three main components that must be available when used on the road.

1) The communication of vehicle ad hoc network: VANET has two-channel communications groups. The safety messages are conducted in the control channel and non-safety messages on the services channel. Vehicles generate ten safety messages every second to other vehicles within a range of 300 meters. These messages aim to help other vehicles stay informed about the situation.

2) Vehicle to Vehicle communication (V2V): The vehicles use three types of communication between them. This type of communication involves the use of a vehicle's computer, which has GPS and radar capabilities. Second is Vehicle-to-Vehicle (V2V) exchange of information [5, 7]. It enables the interchange of speed data, data sharing, and position sharing by allowing cars to communicate with one another and perform its upkeep and secure[6,8].

3) Vehicle to Infrastructure communication (V2I): The third type of communication V2I, which is based on providing data to roadside infrastructure devices such as RSUs. Communication is utilized to deliver customized services like internet access and particular service requests. V2I communication allows cars to request information or services from RSUs and some other roadside infrastructure. For example, a car may inquire about the location of the closest petrol station or cafe, and utilize the RSU may respond. V2I communication may also be utilized to offer internet connection to network devices[5].

The goal of this application was to create a simulation that will perform misbehaving attacks on a certain vehicle. After that, the application will extract the data that it needs, such as the type, speed, and location of the vehicle.

The researchers used a machine learning tool known as WEKA to analyze the data and develop a strategy to identify and prevent misbehavior attacks. We utilized various methods to analyze the collected data, such as the Random tree, the Nave Bayes algorithm, and the Logit Boost method.

Our research aims to know the requirements for achieving security using Vehicular ad hoc network (VANET) and making it a secure and more reliable environment.

1) Develop an application to identify a VANET cybersecurity attack.

2) Improve the Vehicular ad hoc network (VANET) environment to make secure and trust it.

3) A proposed algorithm checks parameters such as vehicle identity sequence of location and speed.

4) Detect the attack and close all vulnerabilities of security holes.

There are a number of cyberattacks that target vehicles on the road, and the VANET environment is not devoid of cyberattacks, like many systems and applications in the world of technology and security challenges.

- There are several questions about this right, the most prominent of which are, what are the examples of the attacks that this environment faces? Among those attacks are Sybil, DDos or Dos attacks, and spam via mail that may affect the work of Vehicular ad hoc network (VANET).
- How to identify these attacks on the Vehicular ad hoc network (VANET) environment which may affect the movement of vehicles, their locations, and the identity of other vehicles?
- Types of attacks that attack the vehicle's identity and location.
- What are the tools used to create a virtual environment for Vehicular ad hoc network (VANET), extract data and analyze the results?

In order to enhance and make this technology safer to use, this paper seeks to analyze cyber security issues and vulnerabilities in the VANET environment, create a safer and more reliable environment, and protect vehicles and their passengers while using it.

II. Literature Reviews

This section seeks to undertake a quantitative analysis of Vehicular ad hoc network (VANET) research conducted between 2007 and 2019. Then, provide an overview of the previous studies in VANET in terms of architecture, simulations, security, and communication protocols focusing on the cybersecurity issues facing VANET [8]. The massive number of publications in the VANET field discusses the various topics related to VANET, such as MAC Layers Issues (MAC-PHY), service, routing, data, tools, mobility, and applications [9].

"NS-2" is the highest level of network simulation software used in the research. Several studies have discussed issues related to the safe and reliable connection of VANETs [10,11].

The vehicle protocol should be reviewed in accordance with the geographical position [12]. Further, the challenges of designing protocol locations based on VANET, including non-DTVANETs, DTVANETs, and hybrids [13,15].

The algorithms mentioned below were created as a result of research and analysis done on the VANET. A Cluster Head (CH), as shown in Fig. 1, is the best option since it optimizes network settings and arranges the structures so that they function with the Adaptive Clustering Protocol (AWCP). Since it provides a protocol for analyzing the movement, position, and speed of the vehicle, the Enhanced Whale Optimization Algorithm (EWOA), one of the modern algorithms, is crucial to the VANET technology.

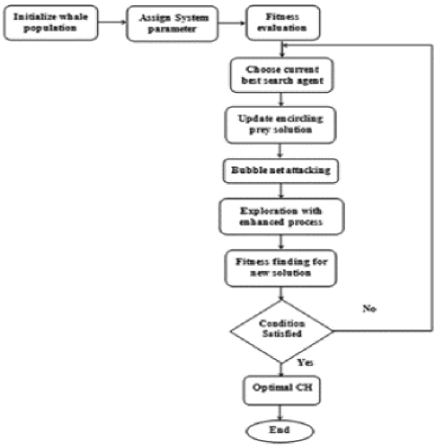

Fig. 1. Flowchart of Optimal CH [18].

Cooper C et al. have conducted a comprehensive study of clustering algorithms designed for VANETs [3]. They proposed classification approaches to address cluster head voting, cluster relationship, and cluster controlling issues, as illustrated in Fig. 2. Furthermore, they evaluated the performance of clustering algorithms and identified the need for practical vehicle channel modeling. Finally, they highlighted the importance of stringent and regulated vehicular channel models.

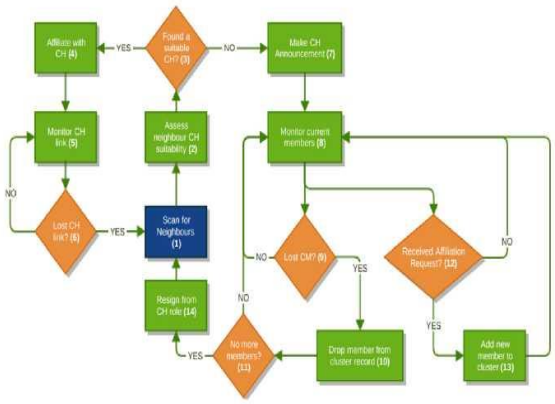

Fig. 2. Flow of a Clustering Algorithm [4].

As a trend, some researchers focused on cyber-attacks and challenges that may face the VANET. For example, R. Engoulou et al. present various architectures and characteristics of VANET. Moreover, they introduced the VANET challenges as time constraints they listed the selection security architecture, requirements, and threads after that. Finally, the authors proposed VANET global security architecture [19].

P. Tyagi et al. examine the routing protocols' features in security and pertinently. Moreover, they proposed an algorithm that focuses on the performance and effort in the most two common protocols used in VANET, named: DSR and AOVD. The protocols mentioned above aim to detect and address a specific type of attack on VANET known as a base black hole attack. The algorithm proposed aimed to enhance AOVD security and detection techniques. Using the proposed algorithm increases the ITS security and reduces the number of malicious nodes.

C. Sowattana et al. showed in Fig. 3 a sample that presents how the Sybil nodes located in the street, propose a distributed technique that detects the Sybil attack by using the messages spread among the nodes. This technique will consider all node positioning locations inside two communication rings as Sybil nodes if it is not acknowledged by one node. [21]

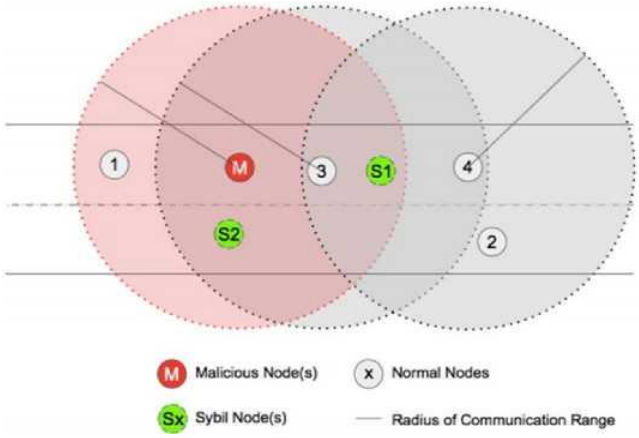

Fig. 3. Sybil attack detection method [20].

M. Hamid et al. [17] present the security issues in VANET communication among the groups. Also, they propose a new solution for man-in-the-middle attacks that may happen in the VANET environment. In [23], authors discuss Invasion of security goals (confidentiality, integrity, availability).

ML is a branch of AI that entails teaching computers to execute tasks and evaluate data. To automate data analysis and processing, ML algorithms employ computer model and decision tools such as decision trees, natural language processing, and neural network models. AI encompasses data exploration and extraction, with machine learning (ML) teaching computers how to use data in decision-making. In this context, data mining is equally significant since it looks for relevant data to execute the task.

ML algorithms are classified into two types: unsupervised and supervised learning. People provide input along with output during training, and the algorithm makes predictions once it has done learning. Unsupervised learning, on the other hand, employs an iterative method nicknamed deep learning which doesn't require the same amount of human input. Unlike supervised learning systems, these algorithms are utilized for more complicated processing tasks. Machine learning techniques are comparable to predictive modeling, but with a concentration on data search [26,28].

This study was conducted with five classifiers from different classification families. It was chosen because they have shown varying ratios in accuracy and timing.

Moreover, the VANET application has found that these classifiers are in line with the results that were found. Five classifiers were used in this study [29-30].

The paper [33] proposes a machine learning-based approach for detecting Sybil attacks in VANETs through collaborative learning. The approach achieves high accuracy in real-time and can be easily integrated into existing VANET

security solutions, highlighting the importance of collaboration between network nodes to enhance security.

The study findings [34] show that existing mechanisms use different detection techniques such as time synchronization, trust-based systems, and clustering algorithms to detect Sybil attacks in VFC environments.

The author in [35] proposes a lattice-based group signature scheme for VANETs that offers forward security and efficient authentication. Through a lattice-based group signature, an efficient and forward-secure authentication protocol can be established for VANETs as is discussed. The proposed scheme is shown to be more efficient than existing schemes in terms of computation time and communication overhead.

A. Summary and Research Gaps

This type of technology is considered a modern type that needs to be studied first, and in these areas, the development is rapid or accelerating. In particular, it may cause dangers and problems that may have unimaginable consequences and lead to the death of some people.

Therefore, it is important for the researcher to work on studying and securing vehicles in order to ensure public security and cyber security in this sector. Additionally, there is a need for more research into secure communication protocols that can be used to protect data exchanged between vehicles. Finally, there is a need for improved methods of detecting and responding to malicious activities in Vehicular ad hoc network (VANET)s. Mapping between the related article and cybersecurity majors is shown in Table I.

TABLE I. MAPPING BETWEEN THE RELATED ARTICLE AND CYBER SECURITY MAJORS

Paper	Finding	Security	Threats	ML/AI	Attack
[14]	Existing VANETs' security concerns and the current strategies for tackling these. They explain how well each solution meets security needs such as identity, integrity, confidentiality, or vehicle revocation.	No	No	Yes	Yes
[16]	Suggested that w identify the Sybil attack in the system using ML employing majority voting.	No	No	Yes	Yes
[18]	suggest an efficient mutual strong authentication for safe Vehicular communications on VANETs	Yes	No	No	Yes
[36]	Proposed framework combines multiple detection sources and achieves a high detection rate with low false positives.	Yes	Yes	No	Yes

III. CYBER SECURITY IN VEHICULAR AD HOC NETWORK (VANET) ATTACKS

The first segment covers four different types of attacks that are designed to interrupt the positioning and identity of vehicles. These attacks can lead to a reduction in the speed of moving vehicles and a decrease in road safety. The misbehaviors that seek to locate or identify cars are covered in the second segment. These can cause a decrease in the comfort level of drivers and the safety of road user [24-25].

A. Experimental Setup on the Simulation.

This section outlines the setups employed in the simulations of the experimentation. The road design is a 10km lengthy roadway with two driving directions and countless tracks in each. The experiment used a total of 12,000 vehicles to replicate. The random speed ranges from 10 km/h to 180 km/h. In addition, the app, VMC, M-VMC, VPMC, or M-VPMC, can be subjected to the misbehavior threats. These designs are adaptable to individual requirements.

B. Examine Simulation Data from VANET Attack

The dataset was created by the VANET information security attack simulation software. The dataset created via simulation has a total of 15594 occurrences. To avoid overfitting, the timestamp, and steps are removed. Each sample contains an ID that can detect the vehicle's identification and position x and y to identify the vehicle's location in the simulation. It is adjustable and dependent on the vehicle's x and y speeds. Furthermore, the type property indicates whether the vehicle is normal, attacker, or virtual. Also, the dataset has been labeled (multi-class) and contains four types of attacks in the column.

C. Types of cyberattacks on vehicles, on the road.

1) Vehicle manipulated coordinates (VMC): An attacker creates or duplicates the ID of the vehicle or several virtual vehicles on the road. This type of attack publishes messages for vehicles with a fake site to a BMS message, so that other vehicles are deluded by the presence of an actual vehicle in the same The left path, the attacker also uses random locations with different time periods.

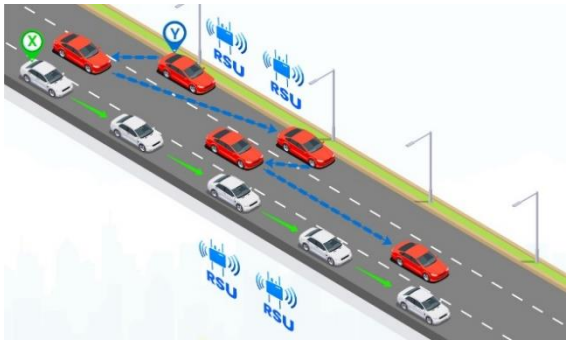

Fig. 4. Vehicle manipulated coordinate.

2) Multi-Vehicle Manipulated Coordinates (M-VMC): The attack in multi-vehicle manipulated coordinates (M-VMC) begins with the creation of a manipulated coordinate's identifier and uses a second vehicle identifier that impersonates a genuine car on the road.

In addition to the fact that this attack works on changing the location of the vehicle in a random manner on the same path as the vehicle on the road, also the same location is used by multiple virtual vehicles for each of them, but they don't use it again, Fig. 5.

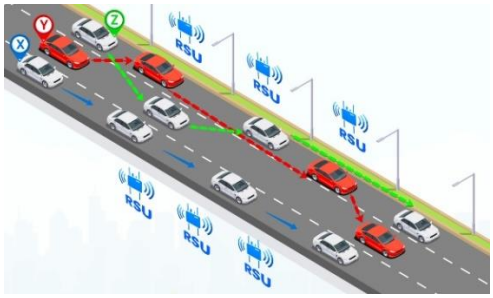

Fig. 5. Multi-vehicle manipulated coordinates.

3) Vehicle Path Manipulated Coordinates (VPMC): A bad type of attack is the vehicle path manipulated coordinates (VPMC) attack. It is an ID-reincarnating virtual vehicle and sends BMS messages as a real vehicle to the vehicles on the road. This type of attack determines its location in advance, similar to a routine traffic.

Fig. 4 to 7, there are two vehicles, X and Y. The real vehicle in red and the arrow is solid, and it is the vehicle X, and its movement path shows the action in different time periods. Stronger, on the other hand, is the virtual vehicle Y in white with a dashed arrow showing its trajectory and virtual movement on the road at different time periods.

In Fig. 6, vehicle X repeats its identification and vehicle Y coordinates are predefined.

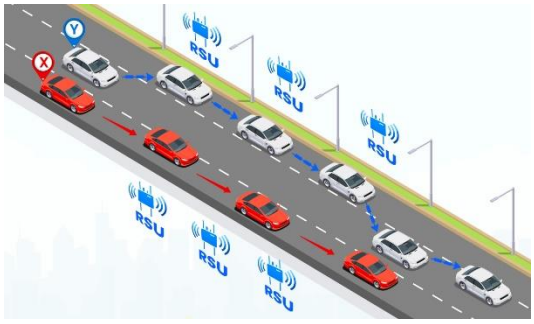

Fig. 6. Vehicle path manipulated coordinates.

a) Multi-Vehicle Path Manipulated Coordinates (M-VPMC): This attack is considered close or similar to VPMC, as it is a bad attack that is difficult to find if it occurs, may cause problems and dangers in traffic safety; and perhaps the most important thing that this attack does is create fake traffic on the road, which causes traffic disruption and delay time Vehicle arrival, The Multi-vehicle path manipulated coordinates (M-VPMC) attacker uses a vehicle and creates broadcast. The attacker can specify the coordinates of the virtual vehicles, bearing to mind that he cannot use a coordinate twice.

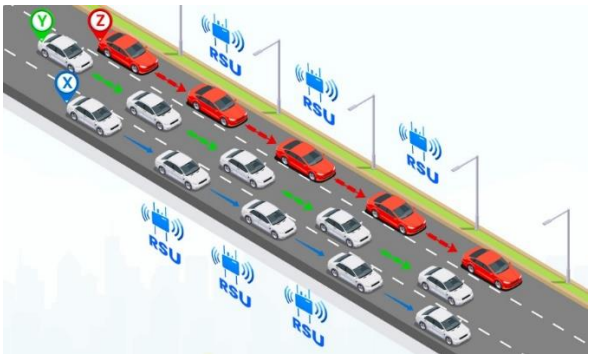

Fig. 7. Multi-vehicle path manipulated coordinates.

D. The Experiment Attack of Vehicular Ad Hoc Network (VANET)

The tools used in this study are discussed in this section. The data collected from the VANET simulation and the experiment are also presented. They are then analyzed using a machine learning tool.

E. The Experiment and Configuration of Simulation

The details of the road configuration used in the experiment are presented in this section. The simulation was conducted on the highway spans a 10-kilometer area divided into two road paths with multiple tracks. The random speed of the 12,000 vehicles that were used during the experiment was set at between 10 and 180 kilometers per hour.

F. The Machine Learning Implementation.

The study was carried out using WEKA, an easy-to-use machine learning tool [26]. It does not require any proficiency in math or programming. In addition, it has a library that can be utilized by developers.

G. Classification Algorithms

The study utilized a classification algorithm that can handle different kinds of circumstances. Some of the systems featured in the experiment are Logit Boost, J48, Naïve Bayes, and Bagging, among others [27-28]. The training phase involves creating a model that will be used on a set of instances. After the model has been trained, it can then be evaluated to see how it performs.

H. The Dataset

The dataset was created using the VANET software. It simulated an information security attack. The dataset created via simulation has a total of 15594 occurrences. To avoid overfitting, the timestamp, and steps are removed. Each sample contains an ID that can detect the vehicle's identification and position x and y to identify the vehicle's location in the simulation. It is adjustable and dependent on the vehicle's x and y speeds. Furthermore, the type property indicates whether the vehicle is normal, attacker, or virtual. Furthermore, the dataset has been labeled (multi-class) and contains four types of attacks in the column.

IV. Result and Analysis

This section aims to talk about the attack described in Section III. It will also discuss the classification tools and the generated dataset.

A. Nave Bayes Output

The first classifier discussed is Naive Bayes, with TP rate representing true positive rate, FP rate representing false positive rate, precision indicating the percentage of true positive designations among all positive categories, recall indicating the percentage of true positive categorizations among all actual positives, and F-measure indicating the normalized harmonic mean of precision and recall. While the Table II, includes these metrics for each class as well as a weighted average of all classes, it does not explain how the statistics were calculated or what they indicate.

TABLE II. Result of Naïve Bayes

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
None	100.00%	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
M-VPMC	53.70%	15.10%	32.40%	53.70%	40.40%	31.40%	86.40%	33.80%
VPMC	23.20%	6.60%	30.40%	23.20%	26.30%	18.70%	86.20%	31.50%
M-VMC	22.30%	5.20%	27.00%	22.30%	24.40%	18.70%	85.80%	24.50%
VMC	12.00%	3.10%	24.20%	12.00%	16.10%	12.50%	85.20%	22.60%
Weighted Avg.	73.20%	3.20%	72.80%	73.20%	72.40%	69.80%	94.60%	72.70%

Similarly, for the additional classifier VMC, the model performs best on the "None" class with 100% TP rate and precision, while the other classes have lower performance metrics. The weighted average of all classes shows that the model has a TP rate of 73.20%, precision of 72.80%, and F-measure of 72.40%, indicating decent overall performance (Fig. 8). In conclusion, this accuracy report provides a comprehensive summary of a machine learning model's performance for a multi-class classifier with five classes, highlighting both the strengths and weaknesses of the model in categorizing examples for each class.

B. Random Forest Outcomes

Fig. 9 illustrates the results of the random tree classifier. The proportion of genuine positive predictions in each class is represented by the TP rate, whereas the proportion of false correct cases is represented by the FP rate. Precision is defined as the ratio of true positive predictions to total positive forecasts for each class, whereas recall is defined as the ratio of true positive predictions to total positive occurrences in each class. The role in the quality mean of accuracy and recall is the F-measure. For the "M-VMC" and "VPMC" classes, the Table III, shows for precision and F-measure as these values could not be calculated due to no positive predictions for these classes. The weighted average row shows the average of the metrics weighted by the number of instances in each class.

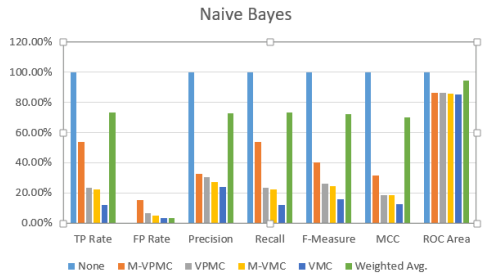

Fig. 8. Result Naïve Bayes.

The classes are listed in the rows of the Table III, with the class name in the leftmost column. The "None" class has the best performance with perfect scores in all metrics. The other classes have varying levels of performance, with M-VPMC having the highest precision but the lowest TP rate and recall, while VMC has the lowest precision but higher TP rate and recall. The weighted average shows that the model has an overall TP rate of 73.0%, precision of 72.5%, and F-measure of 72.4%, indicating decent performance overall.

TABLE III. RESULT OF RANDOM FOREST

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
None	100.00%	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
M-VPMC	43.40%	11.9%	32.9%	43.40%	37.40%	28.00%	85.60%	33.10%
VPMC	35.6%	10.2%	30.40%	35.6%	32.80%	23.80%	85.70%	31.10%
M-VMC	16.5%	4.3%	24.8%	16.50%	19.80%	14.80%	84.47%	22.80%
VMC	13.00%	3.8%	21.7%	13.00%	16.30%	11.70%	84.60%	21.20%
Weighted Avg.	73.00%	3.20%	72.50%	73.00%	72.40%	69.60%	94.30%	72.40%

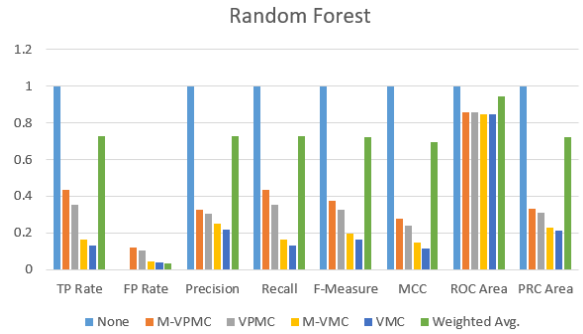

Fig. 9. Result of RF.

C. LogitBoost Classifier Outcomes

The "TP Rate" column (Table IV) displays the true negative rate, which is the number of genuine positives (properly categorized examples) among all positive instances in that class. The "FP Rate" column displays the FPR, which is the amount of false positives (incorrectly categorized occurrences) in comparison to all negative cases in that category. "F-Measure" column represents the F1-score, which is a combined measure of precision and recall. The "Weighted Avg." row at the bottom shows the weighted average of the metrics for all classes.

TABLE IV. OUTPUT OF LOGITBOOST

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
None	100.00%	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
M-VPMC	43.40%	11.9%	32.9%	43.40%	37.40%	28.00%	85.60%	33.10%
VPMC	35.6%	10.2%	30.40%	35.6%	32.80%	23.80%	85.70%	31.10%
M-VMC	16.5%	4.3%	24.8%	16.50%	19.80%	14.80%	84.47%	22.80%

VMC	13.00%	3.8%	21.7%	13.00%	16.30%	11.70%	84.60%	21.20%
Weighted Avg.	73.00%	3.20%	72.50%	73.00%	72.40%	69.60%	94.30%	72.40%

The second row (M-VPMC) has a true positive rate of 0.434, meaning that it correctly classified 43.4% of the instances for this class, and a false positive rate of 0.119, meaning that it incorrectly classified 11.9% of instances as this class when they are actually from other classes. Its precision, recall, and F-measure are 0.329, 0.434, and 0.374 respectively (Fig. 10). Its MCC, ROC area, and PRC area are 0.280, 0.856, and 0.331, respectively.

The following rows show the performance of the model for the other classes (VPMC, M-VMC, and VMC). The last row (weighted average) shows the overall performance of the model, taking into account the number of instances in each class. The model has an overall weighted average true positive rate of 0.730 and a weighted average precision of 0.725, indicating that it performs well overall. The other evaluation metrics such as weighted average false positive rate, recall, F-measure, MCC, ROC area, and PRC area are also reported in the last row [21].

D. The Outcomes of the Bagging Classification

The metrics for evaluating a classification model are shown in the Table V. The model assumes five classes: M-VPMC, VMC, None, M-VPMC, and VPMC. The TP rate is the proportion of actual positive instances properly categorized as positive by the model, whereas the FP rate represents the proportion of actual negative cases wrongly classified as positive by the model. Precision is the percentage of anticipated positive instances that were really positive, whereas recall is the percentage of immediate valid cases that were properly identified as positive (Fig. 11).

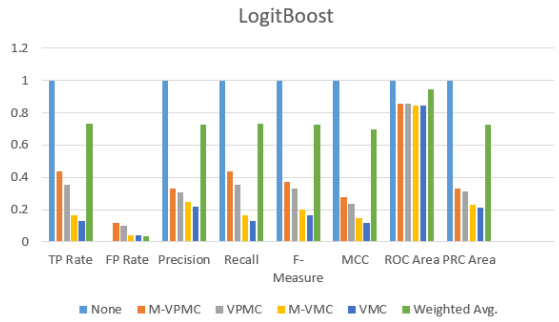

Fig. 10. Output of Logit boost.

TABLE V. OUTCOME OF BAGGING

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
None	100.00%	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
M-VPMC	41.90%	11.5%	32.9%	41.90%	36.80%	27.40%	85.80%	33.50%
VPMC	35.8%	10.0%	31.0%	35.8%	33.20%	24.20%	85.50%	31.70%
M-VMC	15.00%	4.7%	21.6%	15.0%	17.70%	12.20%	84.40%	22.90%
VMC	15.20%	4.1%	23.0%	15.2%	18.30%	13.40%	84.50%	21.30%
Weighted Avg.	72.80%	3.20%	72.40%	72.80%	72.40%	69.50%	94.30%	72.20%

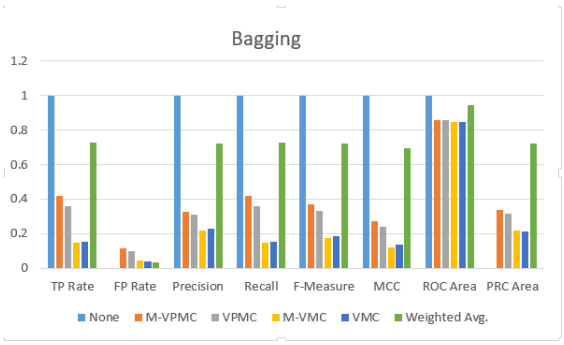

Fig. 11. Output of Bagging.

E. J48 Outcomes

The output of J48 classifier is shown in Fig. 12, presents the performance evaluation metrics for a classification model. For five different classes: M-VPMC, VMC, None, M-VMC, and VPMC. The weighted average of the metrics across all classes is also reported [22].

According to the figure, the model has a greater TP rate and precision for the M-VPMC class than for the other classes, but a lower TP rate and quality for the VPMC class. All the no class has the maximum TP rate and precision, indicating that the model is good at recognizing instances in this class. The total mean metrics(Confusion Metrics) shown in Fig. 13,

indicate that the model performs well overall, with a higher TP rate, accuracy, recall, and F-measure, but a smaller FP rate.

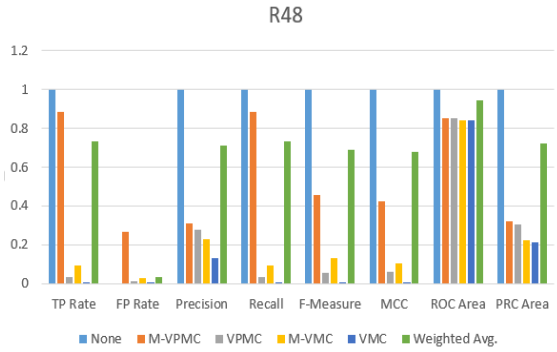

Fig. 12. Output of J48.

```
=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 102  11   0   5   2 |   a = M-VPMC
   6  74   0   5   2 |   b = VMC
   0   0 616   0   0 |   c = None
   9   7   0  86   3 |   d = M-VMC
   9   9   0   3  47 |   e = VPMC
```

Fig. 13. Confusion Matrix for J48

V. Classification and Evaluation

Classification is a well-known ML supervised learning approach that includes predicting a category target variable based on a collection of input characteristics. Generally, the input data is separated into sets for training and testing.

The evaluation process is critical in ML because it determines how well the model is able on unknown data. A classification model's performance may be evaluated using a variety of measures, notably accuracy, precision, recall, F1 score, or ROC curve.

Accuracy: Measures the percentage of correctly predicted instances out of the total instances.

$$\text{Accuracy} = TP + TN / TP + TN + FP + FN$$

Precision: Measures the proportion of true positives out of the total instances classified as positive.

$$\text{Precision} = TP/TP + FP$$

Recall: Recall measures the proportion of true positives out of the total actual positive

$$\text{Recall} = TP/TP + FN$$

F1-Score: The arithmetic mean of accuracy and recall is used to get the F1 score. 2 * (precision * recall) / (precision + recall) is the formula.

A. VANET Attack Detection Flowchart

This section will also talk about the attack detection process. It is shown as a flowchart in Fig. 14.

B. The Generated Dataset

The generated dataset is presented in this section. The simulation program generated a total of 15594 instances of the attack detection process. There were eleven attributes in the dataset, which were sorted into four categories: type, attack type, location, and speed. The steps, timestamp, and detection reason were not included in the analysis. The classification method was also labeled with two classes: normal vehicle and malicious. Fig. 14 shows the flow chart for attack detection.

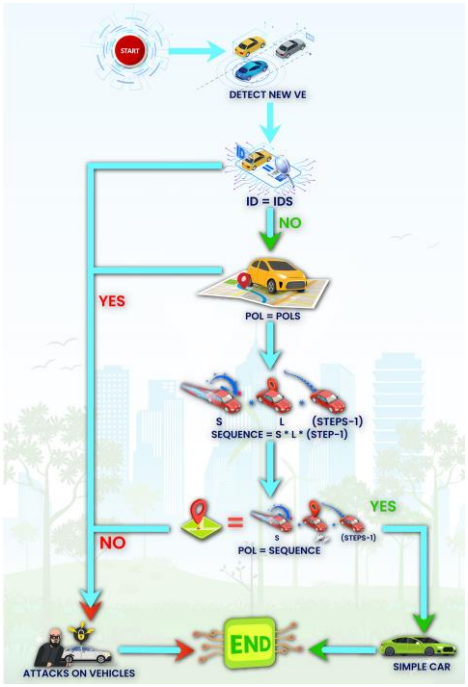

Fig. 14. Flow chart for attack detection.

C. Algorithm for Identifying VANET Attacks

```
Start
    S= start
    L = location
    Ve= vehicles
    PoL= Positions Location
    VPoL= other VPL
    Id = identity
    Vid= other vids
        If (id! =vid)
            If (PoL=! VPoL)
            Sequence=
        S * L * (step-1)
                If (PoL = sequence)
                Output (Normal vehicle)
                Else
                    Output (attacks on it)
                End if
            Else
                Output (attack on it)
            End if
```

D. Detection's Time and Accuracy Performance

The accuracy of five distinct classifiers, including: Bagging, Nave Bayes, Random Tree, J48 & LogitBoost classifiers, is assessed in Fig. 15 based on the total number of examples, correct instances, and wrong instances. The accuracy of cases is used to assess the performance of each classifier. Right examples are those that are correctly identified by the classifier, whereas wrong instances are those that are incorrectly classified. The assessment used a maximum of 263,879 examples, and the findings are shown in detail in the figure below [24]. Instance correctness is an important performance metric since it represents the classifier's accuracy in identifying fresh instances or data points. It represents how accurately the classifier identifies fresh instances or data points, which is an important performance metric. Since Reliability and Accuracy are important factors, it is essential to achieve high instances correctness [25].

Fig. 15. Accuracy of all classifiers.

The given graph depicts the classification performance of five machine learning algorithms: Naïve Bayes, Random Forest, Logitboost, Bagging, and R48, on a dataset. Each algorithm's table displays the number of correctly and incorrectly classified instances. Random Forest attained the highest accuracy, correctly classifying 15593 instances with no misclassification, whereas Naïve Bayes, logitboost, Bagging, and R48 exhibited varying levels of accuracy and misclassification. The selection of the most suitable algorithm depends on the dataset's characteristics, classification objectives, and available computational resources. Choosing a classifier that generates a model promptly is critical, considering both accuracy and efficiency. [26]. Time accuracy graph is shown in Fig. 16.

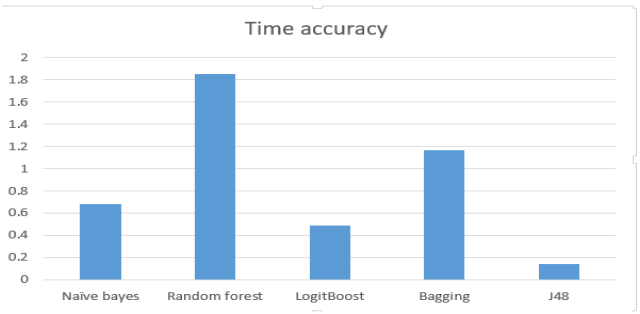

Fig. 16. Time accuracy graph.

VI. Conclusion and Future Work

The present research paper reviews various studies on the security requirements of electronic systems and VANETs (Vehicular Ad Hoc Networks). The paper aims to focus on misbehavior attacks on VANETs, specifically Fake BMS messages that can lead to attacks like M-VPMC, VMC, M-VMC and VPMC. These attacks pose a considerable threat to the security, safety, and comfort of drivers. The study introduces proposed techniques and algorithms to detect misbehavior attacks on identity and location in VANETs. The detection algorithm's performance is evaluated using five classification techniques, including Na¨ıve Bayes, J48, Random tree, Bagging, and LogitBoost, based on accuracy and time taken to build a model/sec. The results reveal that the Random Tree classifier performs the best with accuracy of 99 percentage. The J48 classifier comes in second, with an accuracy of 99.75 and a time of 7.75 sec. On the other hand, the Bagging classifier reports the lowest performance with a time of 15.15 sec and an accuracy of 98.98 percentage.

Using research and scientific studies, along with journals related to VANET, we gained insight into the requirements and security risks that might affect VANET infrastructure. Using a simulation for vehicles and cyber-attacks, we were able to identify whether these threats were the first of four threats posed to this environment in advance. It is very important for the future. It may lead to infrastructure risks and the safety of the driver and pedestrians. In the future, great prosperity is expected for the VANET field, as it is a developed environment that has received a lot of investments at this time. There are still environmental problems and the environment needs to be protected. In the near future, we expect to find many algorithms that will benefit infrastructure owners and vehicle owners. Research has great prospects.

Acknowledgment

The authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work Under Project Number No -PGR-2023-485.

References

[1] H. Alfehaid and S. El Khrdiri, "Cyber security attacks on identity And location of vehicle ad-hoc networks," in Selected Papers from the 12th International Networking Conference: INC 2020 12. Springer, Conference Proceedings, pp. 207–223.

[2] S. Glass, I. Mahgoub, M. J. I. C. S. Rathod, and Tutorials, "Leveraging Manet-based cooperative cache discovery techniques in vanets: A survey And analysis," vol. 19, no. 4, pp. 2640–2661, 2017.

[3] C. Cooper, D. Franklin, M. Ros, F. Safaei, M. J. I. C. S. Abolhasan, And Tutorials, "A comparative survey of vanet clustering techniques," Vol. 19, no. 1, pp. 657–681, 2016.

[4] C. J. Z. S. J. I. I. o. T. J. Guerrero-Ibanez, "Ja internet of vehicles: Architecture, protocols, and security," vol. 5, no. 5, p. 3701, 2017.

[5] R. Atallah, M. Khabbaz, and C. J. I. T. o. V. T. Assi, "Multihop V2i communications: A feasibility study, modeling, and performance Analysis," vol. 66, no. 3, pp. 2801–2810, 2016.

[6] D. Lin, J. Kang, A. Squicciarini, Y. Wu, S. Gurung, and O. J. I. T. o. M. C. Tonguz, "Mozo: A moving zone based routing protocol using Pure v2v communication in vanets," vol. 16, no. 5, pp. 1357–1370, 2016.

[7] H. Ye, G. Y. Li, and B.-H. F. J. I. T. o. V. T. Juang, "Deep reinforcement Learning based resource allocation for v2v communications," vol. 68, No. 4, pp. 3163–3173, 2019.

[8] S. S. Manvi and S. Tangade., "A survey on authentication schemes in Vanets for secured communication, vehicular communications," vol. 9, No. 4, 2017.

[9] Y. Ghasempour, C. R. Da Silva, C. Cordeiro, and E. W. J. I. C. M. Knightly, "Ieee 802.11 ay: Next-generation 60 ghz communication for 100 gb/s wi-fi," vol. 55, no. 12, pp. 186–192, 2017.

[10] K. C. Dey, A. Rayamajhi, M. Chowdhury, P. Bhavsar, and J. J. T. R. P. B. E. T. Martin, "Vehicle-to-vehicle (v2v) and vehicle-to-infrastructure (v2i) communication in a heterogeneous wireless network–performance Evaluation," vol. 68, pp. 168–184, 2016.

[11] A. W. Brittain, A. C. L. Briceno, K. Pazol, L. B. Zapata, E. Decker, J. M. Rollison, N. M. Malcolm, L. M. Romero, and E. H. J. A. j. o. p. m. Koumans, "Youth-friendly family planning services for young people: a systematic review update," vol. 55, no. 5, pp. 725–735, 2018.

[12] T. Yang, Y. Zhang, J. Tan, and T. Z. Qiu, "Research on forward collision Warning system based on connected vehicle v2v communication," in 2019 5th International Conference on Transportation Information and Safety (ICTIS). IEEE, Conference Proceedings, pp. 1174–1181.

[13] R. Oliveira, C. Montez, A. Boukerche, and M. S. J. A. H. N. Wangham, "Reliable data dissemination protocol for vanet traffic safety applications," vol. 63, pp. 30–44, 2017.

[14] F. D. Da Cunha, A. Boukerche, L. Villas, A. C. Viana, and A. A. Loureiro, "Data communication in vanets: a survey, challenges and Applications," 2014.

[15] S. Boussoufa-Lahlah, F. Semchedine, and L. J. V. C. Bouallouche Medjkoune, "Geographic routing protocols for vehicular ad hoc net works (vanets): A survey," vol. 11, pp. 20–31, 2018.

[16] A. S. C. B. Hasrouny, H and A. Laouiti., "Vanet security challenges And solutions: A survey. Veh. Commun." Vol. 7, p. 7–20, 2017.

[17] Hameed, A. G., & Mahmoud, M. S. (2022, September). Vehicular Ad-hoc Network (VANET)–A Review. In *2022 Iraqi International Conference on Communication and Information Technologies (IICCIT)* (pp. 367-372). IEEE.

[18] R. G. Engoulou, M. Bella¨ıche, S. Pierre, and A. J. C. C. Quintero, "Vanet security surveys," vol. 44, pp. 1–13, 2014.

[19] P. Tyagi and D. J. E. i. j. Dembla, "Performance analysis and implementation of proposed mechanism for detection and prevention Of security attacks in routing protocols of vehicular ad-hoc network (vanet)," vol. 18, no. 2, pp. 133–139, 2017.

[20] C. Sowattana, W. Viriyasitavat, and A. Khurat, "Distributed consensus based sybil nodes detection in vanets," in 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, Conference Proceedings, pp. 1–6.

[21] M. N. Mejri, N. Achir, and M. Hamdi, "A new group diffie-hellman Key generation proposal for secure vanet communications," in 2016 13th IEEE annual consumer communications networking conference (CCNC). IEEE, Conference Proceedings, pp. 992–995.

[22] S. Ahmed, M. U. Rehman, A. Ishtiaq, S. Khan, A. Ali, and S. J. J. o. S. Begum, "Vansec: Attack-resistant vanet security algorithm in terms of Trust computation error and normalized routing overhead," vol. 2018, 2018.

[23] I. A. Sumra, H. B. Hasbullah, and J.-l. B. AbManan, "Attacks on Security goals (confidentiality, integrity, availability) in vanet: a survey," In Vehicular Ad-hoc Networks for Smart Cities: First International Workshop, 2014. Springer, Conference Proceedings, pp. 51–61.

[24] C. Wan, J. J. J. o. A. I. Zhang, and H. Computing, "Efficient identitybased data transmission for vanet," vol. 9, pp. 1861–1871, 2018.

[25] P. Asthana, P. Hazela, Bramah %J Multimedia Big Data Computing for IoT Applications: Concepts, and Solutions, "Applications of machine Learning in improving learning environment," pp. 417–433, 2020.

[26] Z. Ge, Z. Song, S. X. Ding, and B. J. I. A. Huang, "Data mining and Analytics in the process industry: The role of machine learning," vol. 5, pp. 20 590–20 616, 2017.

[27] X.-D. Zhang, "A matrix algebra approach to artificial intelligence," 2020.

[28] R. Das and P. M. Khilar, "Driver behaviour profiling in vanets: Comparison of ensemble machine learning techniques," in 2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP). IEEE, Conference Proceedings, pp. 1–5.

[29] L. Nishani and M. J. J. o. I. I. S. Biba, "Machine learning for intrusion Detection in manet: a state-of-the-art survey," vol. 46, pp. 391–407, 2016.

[30] I. Rish, "An empirical study of the naïve bayes classifier," in IJCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, Conference Proceedings, pp. 41–46.

[31] G. Soni and K. Chandravanshi, A Novel Privacy-Preserving and Denser Traffic Management System in 6G-VANET Routing Against Black Hole Attack. Springer, 2022, pp. 649–663.

[32] B. Zhang, X. Wang, R. Xie, C. Li, H. Zhang, and F. J. F. G. C. S. Jiang, "A reputation mechanism based deep reinforcement learning and Blockchain to suppress selfish node attack motivation in vehicular ad hoc network," vol. 139, pp. 17–28, 2023.

[33] S. Azam, M. Bibi, R. Riaz, S. S. Rizvi, and S. J. J. S. Kwon, "Collaborative learning based sybil attack detection in vehicular ad-hoc Networks (vanets)," vol. 22, no. 18, p. 6934, 2022.

[34] H. Yang, Y. Zhong, B. Yang, Y. Yang, Z. Xu, L. Wang, and Y. Zhang, "An overview of sybil attack detection mechanisms in vfc," in 2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W). IEEE, Conference Proceedings, pp. 117–122.

[35] Y. Cao, S. Xu, X. Chen, Y. He, and S. J. C. N. Jiang, "A forwardsecure and efficient authentication protocol through lattice-based group Signature in vanets scenarios," vol. 214, p. 109149, 2022.

[36] D. Ganakwar, "Face detection using logit boost algorithm with ycbcr color space," IJRASET, Vol. 8, no. 1, pp. 184–189, 2020.