Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 27 (2024), No. 7, pp. 2005–2014
DOI : 10.47974/JDMSC-2075

Enhanced security mechanism in vehicular networks using ensemble machine learning to detect malicious activity in VANETs

Atul B. Kathole [@]
Kapil Vhatkar [†]
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology
Pune 411018
Maharashtra
India

Swapnaja Amol Ubale [§]
Department of Information Technology
Marathwada Mitra Mandal College of Engineering
Pune
Maharashtra
India

Vinod V. Kimbahune [‡]
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology
Pune 411018
Maharashtra
India

Amol Dhumane *
Ankur Goyal [#]
Department of Computer Science & Engineering
Symbiosis International Deemed University
Symbiosis Institute of Technology
Pune 411016
Maharashtra
India

[@] *E-mail:* atul.kathole1910@gmail.com
[†] *E-mail:* kapilnv@gmail.com
[§] *E-mail:* swapnaja.b.more@gmail.com
[‡] *E-mail:* vinod.kimbahune@dypvp.edu.in
[*] *E-mail:* amol.dhumane@sitpune.edu.in (Corresponding Author)
[#] *E-mail:* ankur_gg5781@yahoo.co.in

Abstract

One Vehicular Ad-Hoc Network (VANET) consists of numerous active and static vehicles connected in a wireless network. It is an easy and affordable method for transferring information related to traffic and vehicles to the traffic control rooms but may cause the security issues. VANET uses certain protocols for securely transmitting the information and providing internet connectivity to vehicle nodes. An AODV is commonly used in VANET. It is a fast and low-processing machine language model with memory overhead. Vehicles are equipped with OBU, which executes transferring message. If numerous vehicles send their information related to the vehicle simultaneously, then the information gets lost due to collision. Yet, the functions of nodes change frequently, which causes difficulty in routing due to software or hardware default in vehicle. To overcome the abovementioned issues, detection of new attack in VANET is designed. Initially, the data are gathered from online sources. To eliminate unnecessary features from data, data cleaning is performed. To organize the data, the normalization technique is applied. From this data, optimal weighted features are selected using Modernized Random Parameter-based Green Anaconda Optimization (MRP-GAO). The attack detection is performed using the developed Ensemble Machine Learning Model (EMLM) for providing security in VANET network. It is formed by combining the Multi-Layer Perception, SVM, AdaBoost and Bayesian network. Then the fuzzy ranking method is employed to classify the types of attack in VANET. Finally, the performance of the developed secure machine learning-based attack detection in VANET is validated by comparing it with existing techniques and algorithms.

Subject Classification: *Primary 68M12, Secondary 90B18.*

Keywords: *Vehicular ad-hoc network, Security, Modernized random parameter-based green anaconda optimization, Attack.*

1. Introduction

Internet of Things has become the integral part of human life [1]. IoT umbrella holds multiple technology under it including WSN and VANET. In VANET the nodes are interconnected with Road Side and On-Board Units also called as RSU and OBU [2]. RSU is a wireless communication device that collects information related to vehicles along with safety warning and traffic signal and transmits this information to the traffic management system. OBU [3] is similar to an electronic device, which is connected to the vehicle that records the data related to traffic and driving, and then it transmits the data to the satellite navigation system. RSU is a protected framework for receiving information from vehicles [4]. It is located near the traffic lights, roadside and junction point. Intelligent Transport System (ITS) is commonly used in various sensing and communication technologies to improve the safety, flexibility and efficiency of vehicles. VANET is advanced in IT'S to progress the

communication technologies. In distributed cloud communication, vehicle act as nodes to provide secure communication between the clouds having high bandwidth by reducing the delay [2]. In real-time traffic monitoring and surveillance system, VANET provide information related to traffic jam and transfer to the traffic control room. In IT'S, all the vehicles are connected to the cloud for communication with high accuracy [5].

Existing deep learning techniques related to malicious attack detection produce certain limitations and affect the scalability, stability, and ability of the system. IDs system model need full-time monitoring to detect the bugs and it generates both false positive and negative values. Experienced trainers are needed for screening the data, and the system model is so expensive [7].

The contributions are specified below.

- As a first step to implement secure machine learning-based malicious activity detection in VANET, a novel ensemble and fuzzy ranking model is proposed for detecting and identifying any malicious activity in the VANET network.
- For generating the weighted features from the cleansed data, both weights and features are optimally identified using the developed MRP-GAO to enhance the efficacy of detecting the malevolent activities.
- To compare the proposed malicious activity detection in VANET the output attainment will be contracted with various traditional algorithms and methods.

The following are summaries of the remaining parts of this work where the developed malicious activity detection model in VANET is explained: Main characteristics and problems of the contours for malicious activity detection are discussed in Section II. Section III provides a brief of proposed approach using ensemble machining learning, normalization of gathered data, and optimizing weighted features for its detection of malicious activity. The result analysis is made in Section IV, while Section V concludes the work.

2. Related Work

In 2022, Kaur and Kakkar [2] have initiated a hybrid deep learning-based Maxout network for classifying the attacks presented in VANET.

The selection and routing process in the network was performed based on the cluster head selection method. The efficiency of the classification process was improved using the feature selection process. Finally, the classification is carried out using the Detection of Malicious Node (DMN) in the network.

In 2022, Banguiet *et al.* [5] have designed a new model for detecting malicious activities in VANET. This model improved the efficiency of IDS that generated the alerts signal when any fault was detected using Random Forest (RF) algorithm. The accuracy and determination of the IDS were improved based on corsets detection, which means detection based on the particular point. The resultant outcomes compared with various traditional algorithms proved that the proposed model increases the detection accuracy compared to classical techniques.

In 2022,Velayudhan *et al.* [6] have proposed a deep learning-based IDS model using Collaborative Learning Based Sybil Attack (CLBSA-DNN). Here, the vehicles were clustered using the K-Harmonic Mean algorithm (MKHM), and the cluster head was selected using Floyd-Warshall Algorithm. In clustering, Cluster Head (CH) selection helped to transfer data and merger data efficiency. If the condition of CH was normal, then the data has been securely transferred to the cloud computation network. This proposed work attained an accuracy that was better compared to existing traditional algorithm methods.

2.1 *Problem statement*

Malicious activities compromise network security. So, it is necessary to detect and stop such activities and attacks. Most of the deep learning-based detection approaches eliminate the necessary information required to classify the attack and increases the attack detection time. In DMN [2] the computation time is less even in more number of vehicles. It is used to reduce the misconduct behaviors inside the network. But, the cost requirement is high. RF [5] efficiently detects malicious activity in a large amount of vehicular data. It provides better accuracy in real-time attack detection. Yet, it requires more training time and it needs more resources and computational powers. DNN [6] is used to improve the stability of the network by connecting the vehicles. It takes less time to encrypt the node data. Still, it is complex to interpret the results, and more data is required to perform. Yet, the prediction results were not accurate because of the raw data. HDFS [8] is utilized to reduce the information loss of the network. It

is used to cache the network traffic. However, it needs more processing power.

3. Proposed Designed MRP-GAO

The implemented Modernized Random Parameter-based Green Anaconda Optimization (MRP-GAO) algorithm is used in malicious activity detection for optimizing the weights and selecting the features to maximize the variances and correlation coefficient of the system [10]. It reduces the time of the decision-making process and allows the users to focus their time on the analyses. This GAO easily falls into an overfitting problem and uses numerous packets of data for training that leads to loss of some useful information. Therefore, a new optimization algorithm MRP-GAO is developed by randomly updating the best position using the Eq. (1).

$$R_1 = \frac{(BeF_1 \wedge 1.5)}{WoF_{\backprime 1}} \tag{1}$$

Here, the term R_1 is denoted as the random variables and the terms BeF_1, and $WoF_{\backprime 1}$ are the best fitness value and the worst fitness value correspondingly. This improved algorithm (1) automatically repeats the task to boost efficiency and trim down the error.

GAO [6]: It is a population-based metaheuristic algorithm in which green anaconda is considered as the population member [4]. The position of each anaconda is the solution to the problem, where the position in the search space determines the values. The initial position of the members in the search space is randomly selected, and it is represented in Eq. (2).

$$K_{g,h} = an_d + t_{g,h} \cdot (in_f - an_f), g = 1,2...H, h = 1,2,...J \tag{2}$$

Here, the term $K_{g,h}$ is denoted as the h^{th} dimension of the variables, H is represented as the number of anacondas and J is denoted as the number of variables. The term $t_{g,h}$ is denoted as the random variables selected in an interval range [0, 1]. The terms in_f and an_f are denoted as upper bond and lower bond respectively.

3.1 *Collecting Unfiltered Data*

Below is an explanation of the online database that is used by the VANET malicious activity detection framework that was developed. First Dataset (Intrusion Detection Assessment Dataset): The raw data was retrieved from a Kaggle dataset on 2023-07-14 and can be found at https://

www.kaggle.com/datasets/cicdataset/cicids2017. Recent common assaults based on real-world data are included in this dataset [4]. During rush hour, it also has traffic signals that may be time-stamped. Information gathered falls into eleven categories, including traffic signals, network settings, interaction signals, feature set, labeled datasets, accessible protocols, and so on [11].

Algorithm 1: Proposed MRP-GAO

Set the size of the population as KL and the maximum number of iteration WQ

Create the initial population value $GH_f\left(f = 1,2,3,...M\right)$

Calculate the concentration function of the candidates.

Calculate the cumulative function of each of the candidate.

Best position is updated by randomly selecting the candidate.

For $t \to 1toM$

For $Q \to 1toN_p$

$$rand = \frac{(befit \wedge 1.5)}{worsfit}$$

The best candidate solution is updated.

End for

End for

End

4. Experimental Setup

Malicious activity detection has been tackled by the newly-developed Modernized Random Parameter-based Green Anaconda Optimization (MRP-GAO-EMLM) with the help of an Ensemble Machine Learning Model (EMLM). Various classic procedures and methodologies were used to compare the obtained outcomes before execution [12]. The maximum number of iterations for this model was 50, with chromosomal length set at and a total of 10 populations. The results of the experiments were compared with the results of using different algorithms and strategies for detecting harmful activity in VANET [12].

4.1 *Confusion Matrix Assessment*

The assessment of the confusion matrix is to estimate the performance of developed MRP-GAO-EMLM-based model [9], and the analysis is

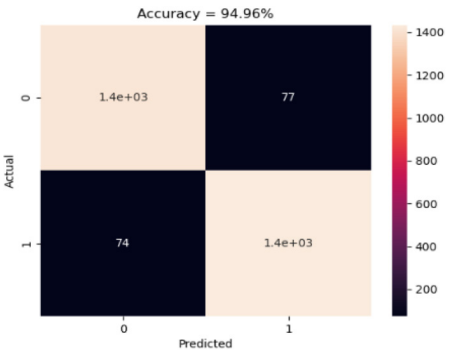

Figure 1

Confusion matrix analysis of the proposed malicious activity discovery with ensemble learning model in VANET

illustrated in Figure 1. From the analysis verification, the developed model attained high prediction in detecting the malicious activities [9].

4.2 *A comparison of the efficiency of various methods and programs*

Figure 2 shows the results of a performance evaluation of the proposed MRP-GAO-EMLM-based [13] VANET malicious activity detection method compared to other conventional algorithms.

We compare the results from various algorithms and strategies to see how well they work. The established model outperformed GAO-EMLM by 11.9%, DA-EMLM by 3.2%, EOA-EMLM by 9.3%, and GAO-EMLM by 6.8% at a learning rate of 0.31. When compared to positive measures, the suggested model obtains a high level of prediction when it comes to detecting hostile activities in VANET [12].

4.3 *Statistical Analysis of Algorithms*

Table 1 displays the statistical performance of the built VANET harmful activity detection system. It strikes a balance between the initiated model's performance and a set of constant statistical measures [13]. When it comes to identifying recently arriving harmful activity in VANET, the initiated model is more efficient than other techniques.

5. Conclusion

Using an ensemble machine learning approach, this malicious activity detection in VANET has been put into practice. The effectiveness

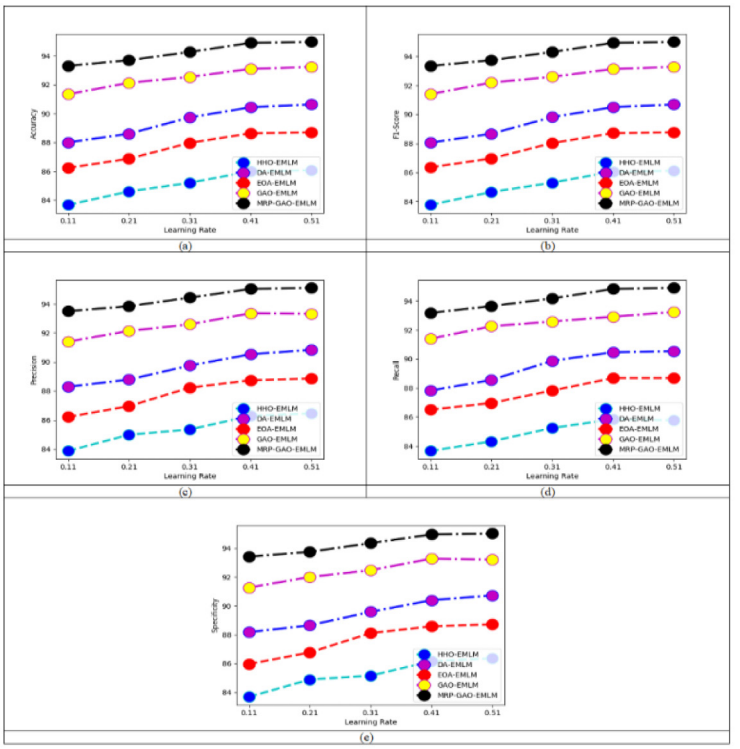

Figure 2

Evaluation of the suggested approach for discovering harmful activities in VANET using a machine learning model

Table 1

Statistical analysis of developed malicious activity detection in VANET among different algorithms

Performance measures	HHO-EMLM [27]	DA-EMLM [28]	EOA-EMLM [29]	GAO-ELM [30]	MRP-GAO-EMLM
Worst	5.407275	5.327993	3.621738	5.916314	5.399695
Best	3.381279	3.38835	3.349509	3.381317	3.219983
Mean	3.587162	3.495432	3.519187	3.474848	3.389149
Median	3.381279	3.38835	3.594434	3.381317	3.219983
Standard Deviation	0.41347	0.322983	0.106858	0.452899	0.547015

of communication between the vehicles is increased by these suggested models. Online sources provide the raw data that are gathered and used in the data cleaning procedure. To maximize the variances and correlation function, the features were chosen and the weights were optimized using the Modernized Random Parameter-based Green Anaconda Optimization

(MRP-GAO). Four classifiers were employed in the classification step once the optimum weighted features were provided. Ultimately, a fuzzy-based ranking technique [14] was used to combine the outputs from four distinct classifiers. The suggested scheme's efficacy in detecting attacks was confirmed by contrasting the obtained results with different methodologies and algorithms. From the Table analysis, the suggested harmful activity detection model outperformed MLP by 6.03%, SVM by 4.05%, and Adaboost by 2.06% in terms of accuracy. In terms of accuracy, f1-score, precision, and recall, the started harmful activity detection model outperformed other existing models.

6. Future Work

Despite the potential benefits of the suggested method, there are still many open questions about how to make VANET malicious activity detection even better. To make data exchange in VANETs more secure and transparent, investigate on how ensemble machine learning and blockchain technology can work together. This could lead to decentralized and tamper-proof event tracking that is relevant to security. The system can optimize its performance and respond to new threats in real-time by creating methods to dynamically adjust the ensemble's composition according to the current state of the network.

References

[1] Amol Dhumane, Rajesh Prasad, and Jayashree Prasad, "Routing Issues in Internet of Things: A Survey," in Proceedings of the International Multi Conference of Engineers and Computer Scientists, Hong Kong, China, pp. 16–18, Mar. 16–18 (2016).

[2] Gurjot Kaur and Deepti Kakkar, "Ad Hoc Networks Hybrid Optimization Enabled Trust-Based Secure Routing with Deep Learning-Based Attack Detection in VANET," Ad Hoc Networks, vol. 136, no. February, Elsevier B.V. (2022).

[3] Ankit Kumar, Ramesh Verma, and Priya Singh, "An improved quantum key distribution protocol for verification," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 22, no. 4, pp. 491–498 (2019).

[4] B. A. Tosunoglu and C. Kocak, "Feature Selection for Clustering and Classification Based Attack Detection Systems in Vehicular Ad-Hoc Networks," *Microprocessors and Microsystems*, vol. 41, no. 2, pp. 1–10, Feb. (2023).

[5] Amol Dhumane, Rajesh Prasad, and Jayashree Prasad, "Routing Issues in Internet of Things: A Survey," in Proceedings of the International Multi Conference of Engineers and Computer Scientists, Hong Kong, China, pp. 16–18, Mar. 16–18 (2016).

[6] Gurjot Kaur and Deepti Kakkar, "Ad Hoc Networks Hybrid Optimization Enabled Trust-Based Secure Routing with Deep Learning-Based Attack Detection in VANET," Ad Hoc Networks, vol. 136, no. February, Elsevier B.V. (2022).

[7] Ankit Kumar, Ankit Sharma, and Priya Mehta, "An improved quantum key distribution protocol for verification," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 22, no. 4, pp. 491–498 (2019).

[8] B. A. Tosunoglu and C. Kocak, "Feature Selection for Clustering and Classification Based Attack Detection Systems in Vehicular Ad-Hoc Networks," *Microprocessors and Microsystems*, Feb. (2023).

[9] Hind Bangui, Rami Dabbous, and Sarah Abdallah, "A Hybrid Machine Learning Model for Intrusion Detection in VANET," Computing, vol. 104, no. 3, pp. 503–31, Mar. (2022).

[10] Nitha C. Velayudhan, Deepak Sharma, and Ravi Kumar, "Sybil Attack Detection and Secure Data Transmission in VANET Using CME-HA-DNN and MD5-ECC," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 2, pp. 1297–309, Feb. (2023).

[11] Ayoub Alsarhan, Fatima Alzahrani, and Omar Khalil, "Machine Learning-Driven Optimization for SVM-Based Intrusion Detection System in Vehicular Ad Hoc Networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 5, pp. 6113–22, May (2023).

[12] Ying Gao, Qiang Zhang, and Li Wang, "A Distributed Network Intrusion Detection System for Distributed Denial of Service Attacks in Vehicular Ad Hoc Network," *IEEE Access*, vol. 7, pp. 154560–71 (2019).

[13] Laisen Nie, Jun Zhang, and Yifan Huang, "Spatio-Temporal Network Traffic Estimation and Anomaly Detection Based on Convolutional Neural Network in Vehicular Ad-Hoc Networks," IEEE Access, vol. 6, pp. 40168–76 (2018).

[14] Teferi Getachew Alemayehu and Zebider Ayenew Ageze, "Fuzzy ideals and fuzzy congruences of distributive demi-p-algebras," *Journal of Discrete Mathematical Sciences & Cryptography*, vol. 26, no. 8, pp. 2127–37 (2018).

Received March, 2024