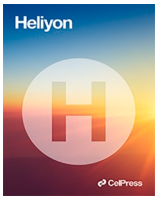
Heliyon 10 (2024) e28844

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Advanced machine learning approach for DoS attack resilience in internet of vehicles security

Nadeem Ahmed [a,*], Fayaz Hassan [a], Khursheed Aurangzeb [b], Arif Hussain Magsi [c], Musaed Alhussein [b]

[a] School of Electronic Science, Beijing University of Posts and Telecommunications, Beijing, 10086, China
[b] Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, P. O. Box 51178, Riyadh, 11543, Saudi Arabia
[c] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 10086, China

ARTICLE INFO

Keywords:
Machine learning
Security
DDoS attack
Vehicular networks

ABSTRACT

Recent years have witnessed security as a great concern in vehicular networks (VANET). Particularly, Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks can jeopardize the network by broadcasting a storm of packets. Correspondingly, the network resources are jammed with malicious traffic. In this connection, the existing research presented various techniques to cope with DoS and DDoS attacks. Different from those traditional approaches, this study proposes an Intelligent Intrusion Detection System (IDS) by leveraging Machine Learning (ML). The proposed IDS utilizes a publicly available dataset on the application layer for mitigating DDoS attacks. The designed ML-based IDS relies on combining both the Random Projection (RP) and Randomized Matrix Factorization (RMF) methods to achieve the best results for enhancing the detection capabilities of the IDS. This amalgamation enhances the system's detection capabilities by extracting and analyzing meaningful features from network traffic data. Experimental validation of our approach involves a comprehensive evaluation of various ML models, including Extra Tree Classifier (ETC), Logistic Regression (LR), and Random Forest (RF). Remarkably, the combined accuracy of these models yields an average system accuracy of 0.98, surpassing existing methods. Unlike conventional approaches, our proposed IDS excels in efficiency and exhibits notable performance in detecting DoS and DDoS attacks in VANET. This proficiency ensures the integrity and safety of vehicle communications. Thus, our research substantially contributes to the vehicular network security field. The presented findings establish a foundation for future advancements in securing connected vehicles.

1. Introduction

The rapid technological advancements in recent years have undoubtedly improved wireless communication, especially in Vehicular Ad hoc Networks (VANETs) [1]. According to CISCO's recent forecast, 66% of the world's population will have access to the internet by the year 2023 and onwards [2]. This corresponds to a global count of 5.3 billion Internet users. Additionally, the report

* Corresponding author.
E-mail addresses: nadeem.ahmed@bupt.edu.cn (N. Ahmed), fayaz.hassan@bupt.edu.cn (F. Hassan), kaurangzeb@ksu.edu.sa (K. Aurangzeb), ahmagsi@bupt.edu.cn (A.H. Magsi), musaed@ccis.ksu.edu.sa (M. Alhussein).

https://doi.org/10.1016/j.heliyon.2024.e28844
Received 10 December 2023; Received in revised form 18 March 2024; Accepted 26 March 2024
Available online 29 March 2024

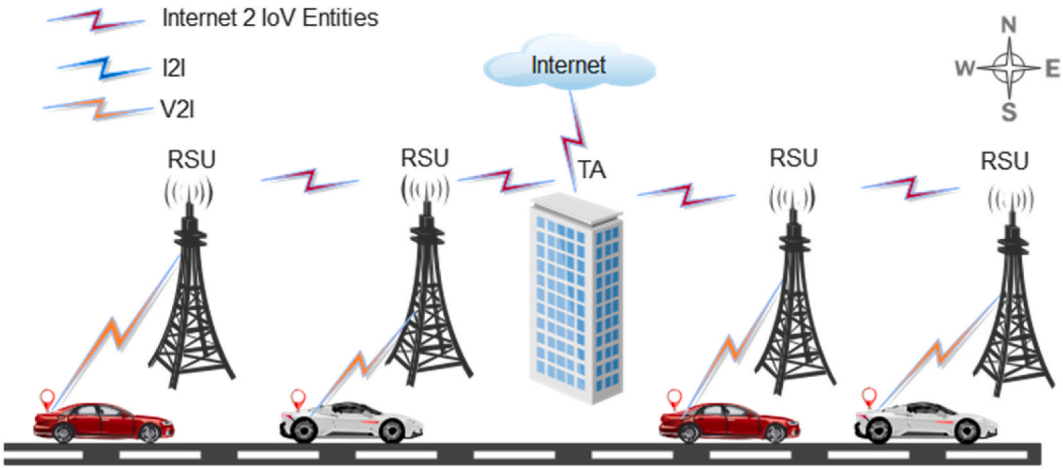


Fig. 1. VANET-based communication prototype.

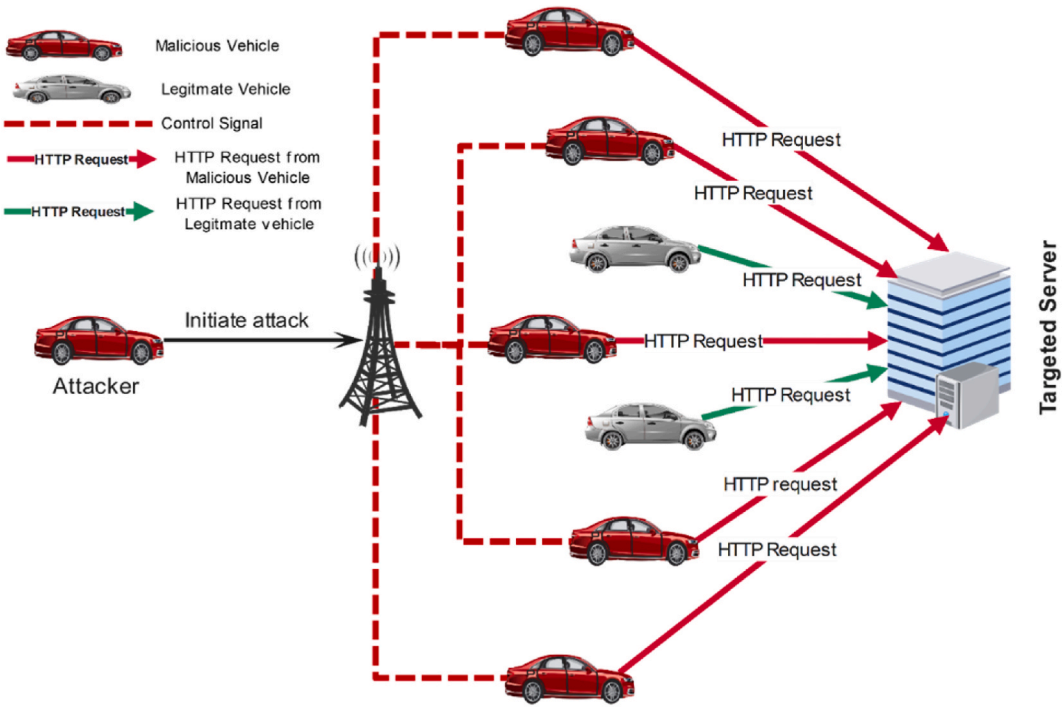


Fig. 2. Illustrative diagram of DoS and DDoS attack in VANET.

foresees a rise in the average number of networked devices per individual, projecting an increase from 2.4 devices per capita in 2018 to 3.6 devices per capita by 2023 and onwards [3]. Similar to mobile phones, home appliances, and televisions, the vehicles are connected in a vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) [4], as illustrated in Fig. 1. VANET connects Transport Authority (TA), Internet, (V2V), and (V2I) through a wireless network, enabling seamless communication for real-time traffic monitoring, data exchange, and enhanced road safety. The VANET-empowered vehicles have powerful processing, communication, and storage capabilities. These vehicles can share various information such as traffic, weather reports, infotainment services, etc. In addition to vehicles, the fixed infrastructure, such as Roadside Units (RSUs) facilitates the VANET to establish vehicle communication. Despite the gigantic features of VANET, such as enhanced road safety, reduced traffic congestion, and improved vehicle communication, a significant challenge is yet to be addressed in ensuring the robust security of vehicles in the face of evolving cyber threats and vulnerabilities. The existing VANET is highly vulnerable to a variety of security threats, such as Denial of Service (DoS) and Distributed DoS (DDoS) attacks [5], Sybil attack [6], Content Poisoning Attack [7], Illusion Attack [8].

The proposed study mitigates DoS attacks in VANET. The DoS attacks are executed by propagating a storm of malicious or repeating

Table 1
List of abbreviations used in the article.

Abbreviation	Meaning
ACL	Access Control List
BOT	Botnet
DNS	Domain Name System
DoS	Denial of Service
DDoS	Distributed Denial of Service
ETC	Extra Trees Classifier
HTTP	Hypertext Transfer Protocol
ICMP	Internet Control Message Protocol
IDS	Intrusion Detection System
IoT	Internet of Things
IoV	Internet of Vehicle
LR	Logistic Regression
ML	Machine Learning
OBU	Onboard Unit
RM	Random Projection
RF	Random Forest
RMF	Randomized Matrix Factorization
SDN	Software-Defined Networking
SYN	Synchronize
TCP	Transmission Control Protocol
TLS	Transport Layer Security
DP	User Datagram Protocol
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
VANET	Vehicular Ad Hoc Network

data [9] to halt or crash the network performance. This storm of malicious data introduces an excessive number of packets from attacker nodes that results in an abnormal traffic. The primary motive of DoS attacks is to compromise the network resources availability, that leads the service interruptions. Redundant packets propagated an attacker node exacerbates delays in VANET services. DDoS attacks, on the other hand, involve multiple coordinated requests aimed at undermining the network's efficiency [10] of the network. In a DDoS scenario, malicious entities can distort the system's availability and services for legitimate users. These attacks disrupt the data flow for legitimate users and exploit various compromised nodes, as illustrated in Fig. 2. The malicious vehicles are observed sending Hyper Text Transfer Protocol (HTTP) requests to victim servers, initiating a flood of network traffic that consumes critical resources such as network speed, bandwidth, and CPU time. Common types of DoS attacks include Domain Name System (DNS) flooding, ping attacks, User Datagram Protocol (UDP) flooding, Internet Control Message Protocol (ICMP) attacks, and Synchronize (SYN) flood broadcast attacks.

The existing literature exploited various techniques to cope with DoS and DDoS attacks, such as identifying similar IP address propagation [11], fuzzy logic-based detection systems [12], and cluster-based attack detection systems [13]. On the other hand, a Machine Learning (ML)-based DDoS attack detection system [14] has been noted in existing literature. Unlike traditional and inadequate solutions, this research leverages an ML-based DOS/DDoS attack detection system at the application layer in VANET.

As in DDoS attacks at the network layer, the attacker uses TCP/UDP and IP spoofing to send several bogus payloads via partially open connections. By sending numerous requests, DDoS attacks overload HTTP and DNS at the application layer [15]. These queries are undetected at the network level and appear valid user requests [16]. Fake requests exceed genuine requests in these cyberattacks. These attacks are difficult to identify because connections and requests from authorized users are already established.

DoS Hulk: DoS Hulk refers to a specific form of DoS attack that specifically targets web servers. This attack strategy targets server with an extensive barrage of HTTP GET or POST requests. The motive of the DoS Hulk is to exhaust the available resources of server that includes CPU, memory, power, and network bandwidth [43,44]. In the result, the network can no longer respond to requests from users. This attack typically introduces vulnerabilities in the server using HTTP requests.

DoS Slowloris: It is a variant of the DoS attack that disrupts the network resources of a targeted web server. Different from DoS Hulk, this attack technique capitalizes on web servers' intricate resource allocation mechanisms. It initiates several partial connections with the server and deliberately maintains these connections in an incomplete state for an extended duration [43,45]. By utilizing this technique, the attacker node effectively prevents the server from timing out and closing these connections. In the result, the server's finite pool of connection slots becomes over occupied, rendering normal users unable to establish relationships. This persistent denial of service conditions arises due to resource exhaustion caused by the attacker's manipulation of the server's connection handling mechanisms.

In this article, an ML-based system is proposed for detecting DDoS attacks. Detecting these attacks manually is complex, and an intelligent protection system is required. The proposed method is simple yet effective and outperforms existing approaches. The research presents below key contributions:

Design a framework for data analysis at the application layer.

To achieve the best-optimized results, much traditional research is done on the DDOS/DoS attack, which relies on the non-learning system. However, the novelty of this study is that an Extra Tree Classifier (ETC), Random Forest (RF), and Logistic Regression (LR) classifier have been chosen to detect the DDoS/DoS attack in Internet of vehicle (IoV)and VANET.

Table 2
Summary of Related Work and their limitations.

Ref:	Dataset	ML DL technique	Aim/Target	Limitation
[20]	KDD'99 and NSL-KDD	SVM method	This algorithm prevented the mutation of well classified events.	Inadequate vehicle communication massive data integration for SVM training.
[21]	DDoS attack dataset on the application layer	MLP and RF	Identifying DDoS Attacks.	Only two models were used despite focusing on accuracy, so enough findings are not available to determine its importance.
[22]	IoTID20 and UNSW-NB15	GBM, RF, ETC	Ensemble-learning-based malicious traffic detection.	The stack design of the model needs a high computational cost.
[23]	DDOS open-source dataset	KNN, SVM, RF.	To protect the banking sector IoT devices from DoS attacks.	Accuracy must be improved in their approach.
[24]	NSL-KDD dataset	Regression Naive Bayes k-Nearest SVM, Logistic KNN	Software – defined networking (SDN)-based ensemble learning for DDoS detection.	The approach requires deploying additional hardware and software components in the SDN architecture, which may increase the implementation complexity and cost of the solution.
[25]	CICIDS2017	RF, SVM, MLP CNN	To suggest an effective ML-based model for detecting DDoS attacks at the application layer.	The paper didn't detail the feature extraction procedure used to preprocess the input data before being fed into the model. This could be a significant element affecting the model's performance.
[8]	NS 3 simulation and used the NSL-KDD dataset	Decision Tree (DT), SVM, KNN	Hybrid data-driven model	The model can detect only known attacks.
[26]	ISCX 2012	SVM, along with addition recursive feature addition	Network traffic attacks	This model ignores class distribution and works on a binary classification method.

ML-based IDS for VANET intrusion detection utilizing Random Projection (RM) and Randomized Matrix Factorization (RMF) feature selection approaches exhibited better outcomes than earlier ML-based methodologies. An efficient DDoS/DoS attack detection framework is presented in this paper.

A comparison with existing studies is undertaken. This research experiments with several ML models to evaluate performance based on accuracy, recall, precision, and F1 score. With reduced computational complexity, the system detects HTTP flood attacks, stealth DDoS/DoS attacks, and stealth DDoS attacks early. A variety of ML algorithms were used to experiment. There was efficient management of ML training within a reasonable timeframe. Additionally, by reducing computational complexity, the designed model ensures the system can detect various attacks.

The study presented here has been divided into five distinct sections. Section 2 offers the related work previously conducted in the field. Section 3 provides an in-depth explanation of the materials and methods that have been utilized throughout the study, including details on the dataset, ML methods, and the proposed methodology that has been employed. Section 4 of the study discusses the results obtained by implementing the proposed method. Section 5 of the study summarizes the discussion and limitations, and Section 6 provides the conclusion drawn from the study. Table 1 illustrates the list of abbreviations used in this paper.

2. Related work

This section examines the current research on VANETs with particular attention to IDS. A key objective of this study is to identify gaps and limitations in the existing knowledge base and use these findings to develop the proposed research.

With a focus on strengthening the dependability and security of vehicle nodes [16], outlines the usage of an IDS flow to identify malicious or misbehaving behaviors of vehicles in VANET. The authors assessed the suggested approach's performance in a VANET environment compared to other algorithms. The paper provides recent research based on IDS, and open issues are discussed very well. However, the authors did not provide any practical implementation.

The authors [17] presented a mutual authentication method that enhances security in VANET by implementing a forward secrecy approach. The process verifies the shared key through batch normalization and can detect impersonation and forgery attacks. The paper presents perfect schemes for comprehensive defense and forward secrecy-enhanced security; however, comparative validation and real-world testing are lacking. This paper [18] indicates several techniques for detecting and safeguarding against network attacks. One such technique is IDS, which falls under three categories: signature-based, anomaly-based, and hybrid structures. The paper concludes that IDS based on signatures can detect anomalous behavior by comparing events to a database of known attacks.

Conversely, anomaly-based IDS monitors attacks by identifying deviations from the system's normal state based on a pre-built database. In either scenario, an alert is generated if a matching similarity is detected or a variation is observed. The paper conducts a comprehensive survey and taxonomy of different deep learning-based approaches for anomaly-based IDS and discusses unresolved research issues. Overall, This paper provides a sufficient, extensive review and gap identification on IDS. Still, it has certain limitations as it is just based on the DL approach, so the ML-based approaches are not considered. This research work [19] highlights that IDS based on signatures can produce fewer false alarms. However, they struggle to maintain an extensive database of potential attack variations since developing signatures for every attack scenario is challenging.

Meanwhile, anomaly-based detection techniques can detect multiple attack types but require additional evaluation resources. Therefore, there is a trade-off between the effectiveness of detection techniques and the resources necessary. This research showcases

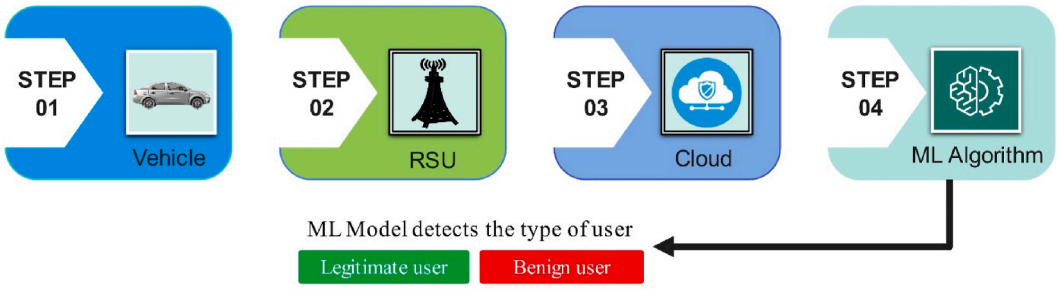


Fig. 3. VANET-based framework: Secure data transmission and threat detection.

remarkable strengths through its comprehensive investigation, seamless integration of fuzzy logic, systematic taxonomy and classification framework, and thorough comparative analysis. Nevertheless, it is essential to acknowledge limitations such as the constrained scope for diverse intrusion patterns, intricate computational requirements linked to fuzzy logic, and the absence of direct focus on real-time operational challenges. Table 2 presents a brief overview of previous studies and their limitations.

In [27], authors used ML techniques to detect DDoS attacks at the application layer, using multi-layer perceptron (MLP) and RF. The results indicated that the RF algorithm attained a high accuracy score 0.999. The MLP algorithm achieved an accuracy score of 0.990 without the big data approach and 0.993 with the big data approach, suggesting that using big data can enhance the MLP algorithm's performance. This study offers key strengths, including addressing Internet of Things (IoT) security, establishing secure boundaries, adapting to evolving threats, and employing multi-technique detection. Nonetheless, limitations arise from dataset dependence and lack of explicit feature selection optimization, impacting generalization and optimization.

Authors in Ref. [28] used statistical methods and ML to identify DDoS attacks within the Software-Defined Networking (SDN) context. The study used a variety of supervised ML algorithms, including LR, Bayesian naive Bayes (BNB), random trees (RT), K-nearest neighbor (KNN), and REPTree. This suggests that the KNN algorithm is a promising approach for detecting DDoS attacks in SDN environments. The novel method addresses DDoS detection limitations in SDNs by introducing a three-section approach collector, entropy-based, and classification-enhancing accuracy. However, potential implementation complexities and resource demands should be evaluated for practical scalability.

In [29], the authors have given a framework capable of detecting malicious traffic of DoS and DDoS attacks on local networks. According to the authors, they have used multi-feature techniques with many classifiers to identify fraudulent traffic. The contribution of this research is promising as the paper presents an ML-based framework for efficient DoS attack detection, demonstrating enhanced model performance across different ML techniques. However, additional investigation into the approach's applicability to various datasets and scalability for real-world deployment is necessary.

The authors in Ref. [30] proposed a security-critical authentication system for the VANET system. This research comprehensively reviews recent and significant intrusion detection methodologies within Vehicular Ad hoc Networks (VANETs), a critical subset of MANETs. The focus on VANET security acknowledges the potentially life-threatening consequences of even a minor security vulnerability. Implementing IDS to safeguard against malicious nodes is pivotal. While the paper presents a comparative analysis of detection techniques and attack focuses, a more detailed examination of the practical applicability of these techniques in real-world VANET security scenarios is a valuable avenue for further exploration.

3. Materials and methods

This section provides a comprehensive overview of this research study framework, which includes a detailed description of the dissemination of attack information among RSUs and vehicles, as illustrated in Fig. 3. To identify malicious or benign users, the framework functions as follows: vehicles transmit data to RSUs, which serve as pathways for data transmission. Following this step, the data is transmitted to a cloud-based architecture, where the designed integrated ML algorithm identifies potentially malicious or benign vehicles and sends the data to the vehicle. This synchronized process facilitates swift threat identification and precise attack classification within the VANET framework.

Next, dataset collection, data preparation procedures, and an analysis of the classifiers used in the research are exploited. Additionally, the model's performance matrices are discussed. As part of our efforts, we aim to ensure transparency and rigor in the research process to make it easier for other researchers to replicate the study and validate its findings. This study aims at establishing a framework that utilizes ML methods to classify and predict DDoS attacks. Several key steps are used; all the steps are illustrated in Fig. 4.

This study was conducted on an Intel Core i9 12th generation PC with 32 GB RAM and a 1 TB SSD running Windows 11 to execute the methodology-based tasks. Also, popular libraries such as TensorFlow and Sci-Kit Learn and a Jupiter notebook to implement the ML methods in Python are used. This research focuses on developing an ML framework that detects DoS and DDoS attacks in VANETs. Once the model is optimized, the resulting output is generated from the model.

3.1. Dataset collection

We have chosen "The ML-Analysis: Application Layer, Dos attack dataset," which examines application-layer Dos and DDoS attacks collected from Kaggle [31]. The dataset contained 78 attributes and 809,361 records. The data was divided into three categories based

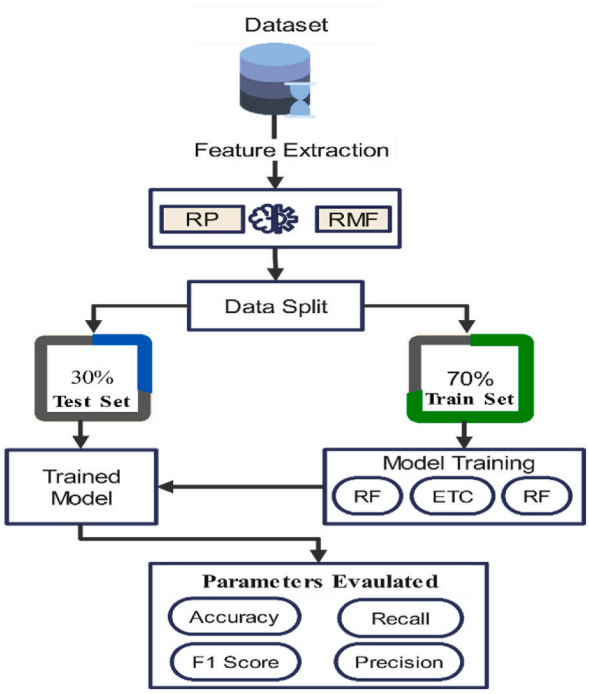


Fig. 4. Flow diagram of the ML framework.

on network analysis: benign is the legitimate traffic, DOS slowloris attacks, and DDOS Hulk attacks. The study aimed to identify and comprehend the properties of these attacks to create efficient countermeasures.

3.2. Data preparation

In Machine learning, data preparation is a crucial aspect that involves identifying significant attributes from the dataset for the training ML model. By selecting relevant features [32], ML models can improve overall performance. After determining the dataset for the ML-based model, we now prepare the dataset that includes feature extraction using RP and RMF; we divided the dataset into training (70%) and testing (30%). This study utilizes feature selection to enhance the learning process to train the machine to achieve the highest accuracy after determining. After the feature selection, the different ML-based classifiers, LR, RF, and ETC are used.

3.2.1. Random projection

In ML, the dimensionality of data plays a pivotal role, influencing both model performance and computational efficiency. High-dimensional datasets often introduce formidable challenges, including the notorious curse of dimensionality and escalated computational demands. In addressing these challenges, Random Projection (RP) emerges as an invaluable technique for dimensionality reduction. RP empowers the transformation of high-dimensional data into a lower-dimensional space while preserving vital data characteristics [33]. It has demonstrated reasonable pairwise distance stability despite its stochastic nature. One of its most notable characteristics is its computational efficiency, enabling large datasets to be analyzed without excessive costs.

RP is expressed in Equation (1).:

$$U = M \times S \tag{1}$$

where:

U represents the reduced feature matrix.

M denotes the original high-dimensional feature matrix.

S is a randomly generated projection matrix.

It is a valuable method for reducing feature space dimensions in machine learning. Using a randomly generated projection matrix S, the original high-dimensional feature matrix M is linearly projected into a lower-dimensional space. The result is matrix U, which reduces dimensionality and captures the essence of the data. Multiple ML applications rely on this process to boost computational efficiency.

3.2.2. Randomized Matrix Factorization

RMF addresses high-dimensional data challenges in machine learning. Additionally, the RMF makes ML algorithms more effective in handling complex datasets due to its computational efficiency, scalability, and feature extraction capabilities [34].

Equation (2) represents the iterative update for X in RMF.

$$X \leftarrow B\left(Y^T\left(YY^T\right)\right)^{-1}\Big) \tag{2}$$

In Equation (2), X is optimized while Y is constant.

Equation (3) represents the iterative update for Y:

$$Y \leftarrow \left(X^T\left(XX^T\right)\right)^{-1}\Big)B \tag{3}$$

In Equation (3), Y is optimized while holding X constant.

3.2.3. Feature extraction optimization using RP and RMF

The RP and RMF are combined in data analysis because they have complementary strengths and can provide more robust results. In the present study, we have selected fifty features, incorporating the top 25 features obtained from each method, and further analysis of the performance is conducted. The selection of fifty features is based on a comprehensive analysis of the dataset's complexity and the trade-off between dimensionality reduction and information retention. Through extensive experimentation, it was observed that the inclusion of the top 25 features from each method (RP and RMF) struck an optimal balance. This choice maximizes the information gain while mitigating the risk of overfitting, ensuring a robust and generalizable model for the subsequent analysis.

3.3. Classification

After data preprocessing, for the ML model, Several performance metrics are used to evaluate the model's performance, including accuracy, precision, recall, and the F1 score. This research uses ML classifiers to classify benign, DoS Hulk, and DoS Slow Loris attacks, and we have used LR, ETC, and RF classifiers to measure the efficiency of identifying different attacks. A brief description of the used classifiers is presented herein.

3.3.1. Extra Tree Classifier

The Extra Trees Classifier is an ensemble learning algorithm for classification tasks [35]. Using training data and features, it constructs a collection of decision trees [36]. Each decision tree in the group is built using randomized feature selection and split points, leading to further variance reduction in the ensemble. A decision boundary from each tree in the collection forms the decision boundary of the ensemble. In making a prediction, the Extra Trees Classifier outputs the class that has received the most votes from the collection of trees. The randomization of both the training data and feature selection contributes to the reduction of overfitting and an increase in the diversity of the trees in the ensemble, which results in improved classification accuracy.

3.3.2. Random Forest

Among ensemble learning algorithms, Random Forest is widely used for classification and regression analysis [37]. The system consists of multiple decision trees, each trained independently using a randomly selected subset of the training data and features. At each split point, the algorithm randomly selects a subset of features to reduce overfitting and improve the diversity of the trees [38]. To form the final prediction, the predictions of all trees in the forest are combined.

Random Forest is particularly effective in handling high-dimensional data and nonlinear relationships between features.

3.3.3. Logistic regression

It is a statistical model used in machine learning and statistical analysis to analyze datasets with binary dependent variables, which can take only two values [39]. In logistic regression, we utilize the modified logistic function to model the probability of a connection between the dependent and independent variables [40]. Equation (4) expresses the modified logistic function is expressed as:

$$P(Y=1) = \frac{e^{\rho}}{1+e^{\rho}} \tag{4}$$

Here, P(Y = 1) represents a probability estimate, and e^{ρ} is the input to the modified logistic function. The modified logistic function is a mathematical formula widely used in machine learning for estimating the likelihood of an event occurring. It maps a curve from negative infinity to positive infinity, with the output bounded between 0 and 1. The function is based on an exponential function and yields an output 0 as the curve approaches negative infinity and 1 otherwise.

Logistic regression is essential for modeling the relationship between explanatory and response variables in binary classification tasks. It provides valuable insights and predictions for various applications.

3.4. Performance metrics of classifier

The following performance metrics evaluate the classifier's performance.

3.4.1. Accuracy

Accuracy is a widely adopted performance metric in ML that quantifies a model's percentage of accurately classified data points

relative to the total number of data points evaluated [41]. Mathematical accuracy can be expressed in Equation (5)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where:

TP represents True Positive.
TN represents True negative.
FP represents a false positive.
FN represents negative.

3.4.2. Recall

Recall is a widely used metric for evaluating the performance of classification models. This metric quantifies the capability of a model to accurately recognize all pertinent instances in a dataset, commonly referred to as positive cases, while simultaneously reducing the number of false negatives [41]. By measuring recall, one can assess how comprehensively a classifier identifies relevant instances while minimizing incorrectly classified negatives. Recall measures the model's ability to identify a dataset's complete set of pertinent samples while reducing the likelihood of missing positive cases. Mathematically, recall is expressed by Equation (6).

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

3.4.3. Precision

Precision is used as a performance metric to evaluate the accuracy of a classifier. This metric evaluates the classifier's capacity to identify relevant instances while minimizing false positives [42]. It is determined by dividing the number of true positives by the sum of true positives and false positives. The mathematical expression for precision, denoted as Equation (7),

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

3.4.4. F1 score

The F1 Score measures the model's accuracy with respect to the dataset [42]. The F1 score, as represented by Equation (8). Assess the model's accuracy in relation to the dataset. It is computed using the equation below,

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

4. Results

In this study, we evaluated the performance of an IDS using an ML model to detect DoS and DoS attacks on VANETs. We applied three ML models to classify benign and malicious outcomes: RF, LR, and ETC. The feature selection process was enhanced by combining RP and RMF. All the classifiers' confusion matrices and tables provide a clear overview of the predicted outcomes based on

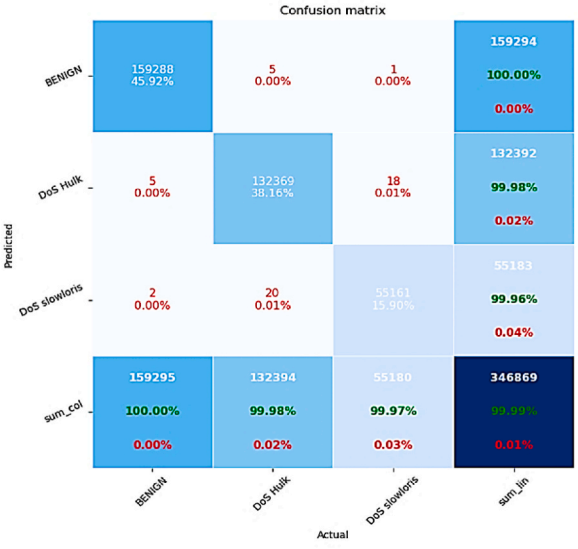


Fig. 5. Confusion matrix of RF classifier.

the designed model. With RP and RMF, we evaluated the performance of the classifiers. The results demonstrated that combining these two feature selection techniques significantly improved the accuracy of the classifiers. Every classifier result is shown in a separate confusion matrix.

RF Classifier: The confusion matrix of the RF classifier shows that RF predicted benign outcomes with 100% accuracy, while the accuracy for DOS Hulk and DOS slowloris was 99.98%, respectively, as shown in Fig. 5. Fig. 6 displays the ROC curve for the RF classifier.

LR Classifier: Fig. 7 displays a confusion matrix for the logistic regression classifier, which summarizes the predicted outcomes. By analyzing the obtained result LR.

In the classifier confusion matrix, the relationship between true positive and false positive, as well as true negative and false negative, the result shows that benign accuracy is 97.64%. In contrast, for DOS Hulk and DOS slowloris, the accuracy is 94.29% and 98.18%, respectively. Fig. 8 shows the achieved ROC curve of LR.

ET Classifier: Fig. 9 displays a confusion matrix for the Extra Tree classifier, which summarizes the predicted outcomes by analyzing the plot. The benign results with 99.9% accuracy, while for DOS Hulk and DOS slowloris, the accuracy is 99.98% and 99.97%, respectively. Fig. 10 shows the ROC curve of the LR classifier of our proposed model.

A combined chart of the ML models exhibits the performance matrices, demonstrating the precision, recall, and score in Fig. 11: Consolidated results of ML models for individual classes.

Table 3 represents the classification report of the ML model with RP and RMF.

Table 4 shows the comparative analysis of existing work with the results we achieved. The table provides the research year, models, and accuracy of different research studies compared to our proposed study. The Accuracy column represents the combined average accuracy of LR, ETC, and RF classifiers, offering an overarching metric for comparison.

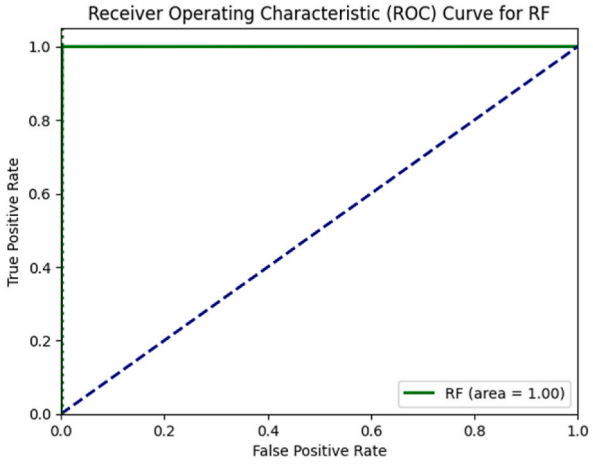


Fig. 6. ROC Curve of RF classifier.

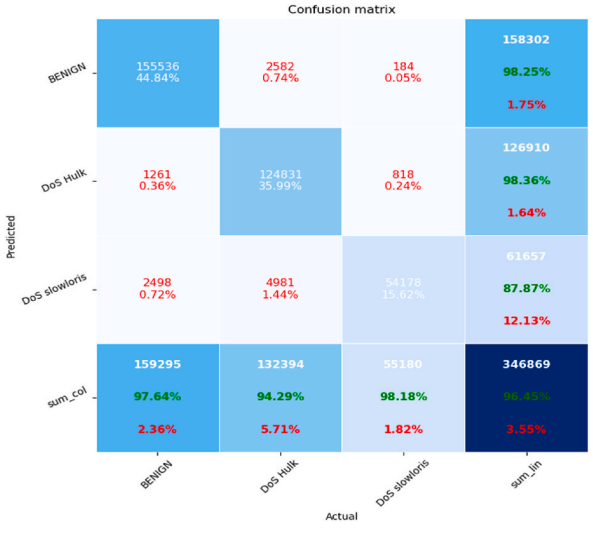


Fig. 7. Confusion matrix of LR classifier.

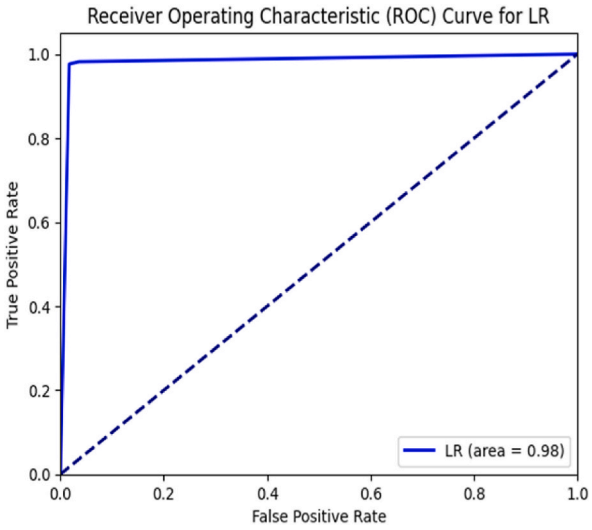


Fig. 8. ROC curve of LR classifier.

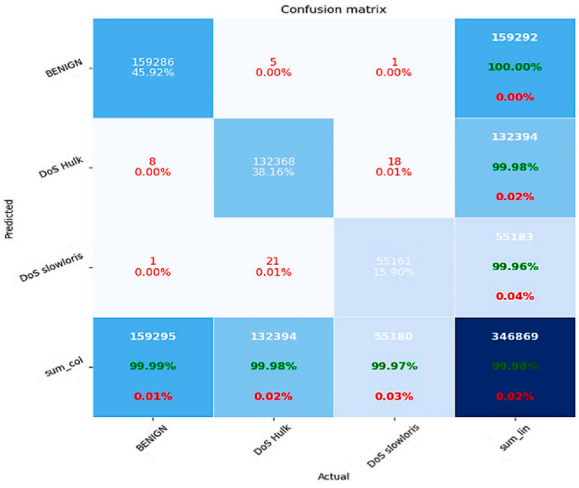


Fig. 9. Confusion Matrix of Extra Tree classifier.

4.1. Discussion and limitations

VANETs continue to experience increased attacks [48–50], including DDoS attacks. There is a need for automated risk mitigation solutions. In this regard, the recommended approach employs ML to detect DoS and DDoS attacks in real time. This study provides a straightforward and efficient way to achieve robust results, making it a viable option for implementing network interfaces to detect attacks, especially those that target application layers, such as HTTP floods that employ HTTP post requests for server access. Compared to non-learning, the technique has significant advantages. It operates in real-time, enabling prompt responses to future attacks. Furthermore, the solution is straightforward to deploy and requires few changes to the existing network infrastructure. In addition, it can be readily adapted to detect attacks on application layers, allowing it to be utilized in various scenarios.

Notwithstanding the study's positive outcomes, several limitations must be considered when interpreting findings. This solution was specifically tested on a single dataset, highlighting the need for extensive studies to see how well it can identify attacks on several layers. Also, to potentially produce better results for ML models, future studies should examine the effect of dataset size on the approach's effectiveness. Finally, our findings suggest a feasible technique for identifying and mitigating DoS/DDoS attacks in VANETs. While there are certain limits, additional study analysis can improve the effectiveness of this research approach, making it a worthwhile addition to the arsenal of instruments for safeguarding vehicular networks.

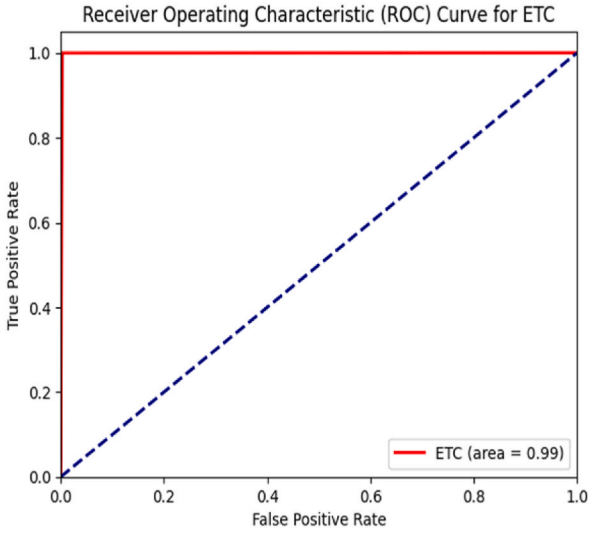


Fig. 10. ROC curve of ET classifier.

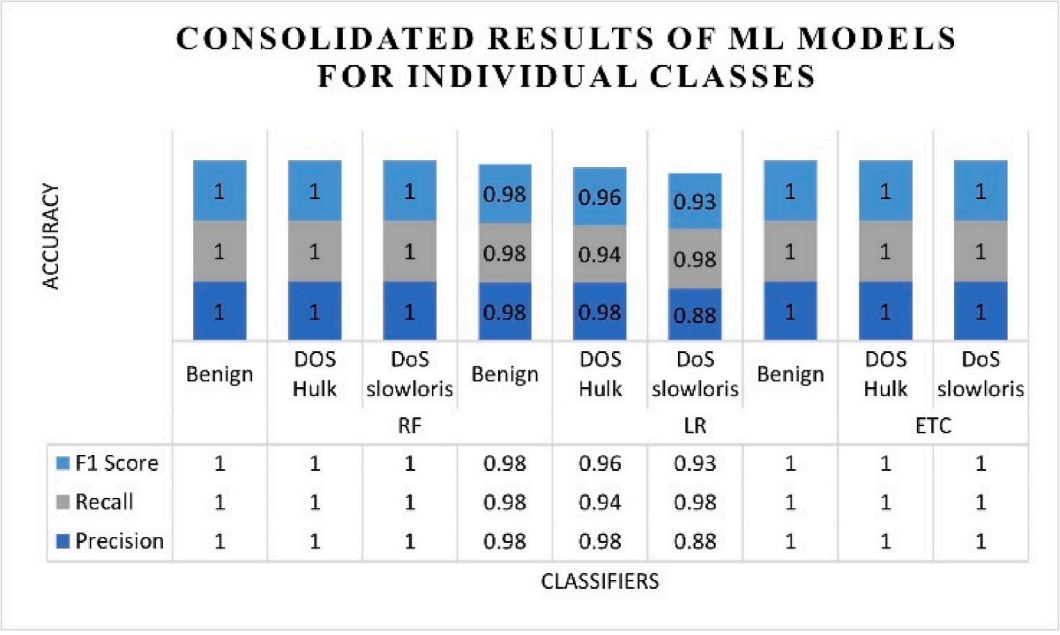


Fig. 11. Consolidated results of ML models for individual classes.

Table 3
An analysis of the classification of ML models using RP and RMF features.

Classifier	Class	Precision	Recall	F1 Score
Random Forest	Benign	1	1	1
	DOS Hulk	1	1	1
	DoS slowloris	1	1	1
Logistic Regression	Benign	0.98	0.98	0.98
	DOS Hulk	0.98	0.94	0.96
	DoS slowloris	0.88	0.98	0.93
Extra Tree classifier	Benign	1	1	1
	DOS Hulk	1	1	1
	DoS slowloris	1	1	1

Table 4
Comparison of classifier accuracies between this research and existing studies.

Research	Model	Research Year	Accuracy
[43]	K-Mean	2022	0.95
[44]	Deep belief Network technique	2021	0.96
[45]	RF, SVM	2021	0.988
[46]	SVM	2022	0.97
[47]	NB	2020	0.985
This research	LR, ETC, RF	2024	0.987

5. Conclusion and future work

Since DoS and DDoS attacks are vulnerable to VANETs, this paper provides an ML-based framework for application layer DoS and DDoS attack detection in VANETs. This framework uses a combination of features extracted with RP and RMF to improve the accuracy of detecting these attacks in our experiments with the classifiers RF, ETC, and LR. The classifier RF obtained the best results with the combined features. These results demonstrate the effectiveness of the proposed framework in improving the security of VANETs against DoS/DDoS attacks. In future work, we may choose the dataset that should be generated on the road networks simulation results and investigate the impact of different datasets with more classifiers, as Deep Learning (DL) classifiers can be evaluated for further investigation. Furthermore, this study is limited to DOS/DDOS attack detection, which can be further assessed regarding energy consumption and computational complexity in future work. The current research is focused on determining the combined RP and RMF methods. Hyperparameter tuning was omitted to underscore the inherent robustness of these techniques. We aimed to emphasize the specific contributions of RP and RMF without introducing other complexities, and we got outstanding results. While recognizing the importance of hyperparameter optimization, our study provides a focused exploration of these methods within the context of vehicular network security. This study contributes to IoV security and lays the foundation for further research. Moreover, this work can be extended by evolving data volume and computational technologies.

Data availability statement

The data used to support the findings of this study are available from the first author upon request.

Credit authorship contribution statement

Nadeem Ahmed: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Fayaz Hassan:** Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Investigation, Formal analysis, Conceptualization. **Khursheed Aurangzeb:** Writing – original draft, Validation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Arif Hussain Magsi:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Musaed Alhussein:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research is funded by Researchers Supporting Project Number (RSPD2024R947), King Saud University, Riyadh, Saudi Arabia.

References

[1] E.C. Eze, S.-J. Zhang, E.-J. Liu, J.C. Eze, Advances in vehicular ad-hoc networks (VANETs): challenges and road-map for future development, Int. J. Autom. Comput. 13 (2016) 1–18.
[2] T. Nguyen, H.-L. Mai, R. Cogranne, G. Doyen, W. Mallouli, L. Nguyen, M. El Aoun, E.M. De Oca, O. Festor, Reliable detection of interest flooding attack in real deployment of named data networking, IEEE Trans. Inf. Forensics Secur. 14 (9) (2019) 2470–2485.
[3] Cisco, "Cisco Annual Internet Report (2018–2023) White Paper," Cisco Executive Perspectives, Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.
[4] H. Khelifi, S. Luo, B. Nour, H. Moungla, Y. Faheem, R. Hussain, A. Ksentini, Named data networking in vehicular ad hoc networks: state-of-the-art and challenges, IEEE Commun. Surv. Tutorials 22 (1) (2019) 320–351.
[5] A.H. Magsi, S.A.H. Mohsan, G. Muhammad, S. Abbasi, A machine learning-based interest flooding attack detection system in vehicular named data networking, Electronics 12 (18) (2023) 3870.
[6] B. Yu, C.-Z. Xu, B. Xiao, Detecting sybil attacks in VANETs, J. Parallel Distr. Comput. 73 (6) (2013) 746–756.
[7] A.H. Magsi, L.V. Yovita, A. Ghulam, G. Muhammad, Z. Ali, A content poisoning attack detection and prevention system in vehicular named data networking, Sustainability 15 (14) (2023) 10931.

[8] A.H. Magsi, G. Muhammad, S. Karim, S. Memon, Z. Ali, Push-based content dissemination and machine learning-oriented illusion attack detection in vehicular named data networking, Comput. Mater. Continua (CMC) 76 (3) (2023).
[9] K. Verma, H. Hasbullah, A. Kumar, Prevention of DoS attacks in VANET, Wireless Pers. Commun. 73 (2013) 95–126. Springer.
[10] A. Gaurav, B.B. Gupta, F.J.G. Peñalvo, N. Nedjah, K. Psannis, DDoS attack detection in vehicular ad-hoc network (VANET) for 5G networks, in: Security and Privacy Preserving for IoT and 5G Networks: Techniques, Challenges, and New Directions, Springer, 2022, pp. 263–278.
[11] S. Kumawat, H. Kaur, O. Dahiya, An analytical study on intrusion detection system in integrated vehicular ad-hoc network attacks, in: Proc. 3rd Int. Conf. Intell. Eng. Manag, ICIEM, 2022, pp. 378–383, 2022.
[12] N. Vanets, An Adaptive Real-Time Malicious Node Detection Framework, 2023.
[13] R. Kolandaisamy, R.M. Noor, I. Kolandaisamy, I. Ahmedy, M.L.M. Kiah, M.E.M. Tamil, T. Nandy, A stream position performance analysis model based on DDoS attack detection for cluster-based routing in VANET, J. Ambient Intell. Hum. Comput. 12 (2021) 6599–6612.
[14] N. Ahmed, Z. Deng, I. Memon, F. Hassan, K.H. Mohammadani, R. Iqbal, A survey on location privacy attacks and prevention deployed with IoT in vehicular networks, Wireless Commun. Mobile Comput. 2022 (2022).
[15] S. Amaouche, S. Benkirane, A. Guezzaz, M. Azrour, A proposed machine learning model for intrusion detection in VANET, Lect. Notes Networks Syst. 635 (LNNS) (2023) 103–108.
[16] F. Gonçalves, J. Macedo, A. Santos, Evaluation of VANET datasets in context of an intrusion detection system, in: 2021 29th Int. Conf. Software, Telecommun. Comput. Networks, SoftCOM, 2021, 2021.
[17] I. Naqvi, A. Chaudhary, A. Rana, Intrusion detection in VANETs, in: 2021 9th Int. Conf. Reliab. Infocom Technol. Optim., Trends Futur. Dir. ICRITO, 2021, 2021.
[18] M. Yao, X. Wang, Q. Gan, Y. Lin, C. Huang, An improved and privacy-preserving mutual authentication scheme with forward secrecy in VANETs, Secur. Commun. Networks 2021 (2021).
[19] A. Aldweesh, A. Derhab, A.Z. Emam, Deep learning approaches for anomaly-based intrusion detection systems: a survey, taxonomy, and open issues, Knowl. Base Syst. 189 (Feb) (2020).
[20] A. Alsarhan, M. Alauthman, E. Alshdaifat, A.R. Al-Ghuwairi, A. Al-Dubai, Machine Learning-driven optimization for SVM-based intrusion detection system in vehicular ad hoc networks, J. Ambient Intell. Hum. Comput. 1 (2021) 1–10. Feb.
[21] M.J. Awan, et al., Real-time DDoS attack detection system using big data approach, Sustain. Times 13 (19) (2021) 10743. Sep. 2021.
[22] P.L. Indrasiri, E. Lee, V. Rupapara, F. Rustam, I. Ashraf, Malicious traffic detection in IoT and local networks using stacked ensemble classifier, Comput. Mater. Continua (CMC) 71 (1) (2021) 489–515. Nov.
[23] U. Islam, et al., Detection of distributed denial of service (DDoS) attacks in IOT based monitoring system of banking sector using machine learning models, Sustain. Times 14 (14) (2022) 8374. Jul. 2022.
[24] N. Ahuja, G. Singal, D. Mukhopadhyay, N. Kumar, Automated DDOS attack detection in software defined networking, J. Netw. Comput. Appl. 187 (2021) 103108. Aug.
[25] M.K. Kareem, O.D. Aborisade, S.A. Onashoga, T. Sutikno, O.M. Olayiwola, Efficient model for detecting application layer distributed denial of service attacks, Bull. Electr. Eng. Informatics 12 (1) (2023) 441–450. Feb.
[26] T. Hamed, R. Dara, S.C. Kremer, Network intrusion detection system based on recursive feature addition and bigram technique, Comput. Secur. 73 (2018) 137–155. Mar.
[27] I. Ahmad, Z. Wan, A. Ahmad, A big data analytics for DDOS attack detection using optimized ensemble framework in Internet of Things, Internet of Things 23 (2023) 100825. Oct.
[28] A. Banitalebi Dehkordi, M.R. Soltanaghaei, F.Z. Boroujeni, The DDoS attacks detection through machine learning and statistical methods in SDN, J. Supercomput. 77 (3) (2021) 2383–2415. Mar.
[29] F. Rustam, M.F. Mushtaq, A. Hamza, M.S. Farooq, A.D. Jurcut, I. Ashraf, Denial of service attack classification using machine learning with multi-features, Electron 11 (22) (2022) 1–20.
[30] A systematic review of the intrusion detection techniques in VANETS, TEM J. 11 (2) (2022) 900–907.
[31] ML Analysis: Application Layer DoS Attack Dataset | Kaggle." [Online]. Available: https://www.kaggle.com/code/hamzasamiullah/ml-analysis-application-layer-dos-attack-dataset/notebook. [Accessed: 30 November,-2023].
[32] S. Bahassine, A. Madani, M. Al-Sarem, M. Kissi, Feature selection using an improved Chi-square for Arabic text classification, J. King Saud Univ. - Comput. Inf. Sci. 32 (2) (2020) 225–231. Feb.
[33] T. Mo, et al., Classifier ensemble with evolutionary optimisation enforced random projections, Expert Syst. Appl. 222 (2023) 119845. Jul.
[34] E. Nasiri, K. Berahmand, Y. Li, Robust graph regularization nonnegative matrix factorization for link prediction in attributed networks, Multimed. Tool. Appl. 82 (3) (2023) 3745–3768. Jan.
[35] A. R. Kharwar and D. V. Thakor, "An Ensemble Approach for Feature Selection and Classification in Intrusion Detection Using Extra-Tree Algorithm," https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJISP.2022010113, vol. 16, no. 1, pp. 1–21, Jan. 1AD.
[36] M. Sarhan, S. Layeghy, M. Portmann, Towards a standard feature set for network intrusion detection system datasets, Mobile Network. Appl. 27 (1) (2022) 357–370. Feb.
[37] S. Wali, I. Khan, Explainable AI and Random Forest Based Reliable Intrusion Detection System, 2021. Dec.
[38] T. Wu, H. Fan, H. Zhu, C. You, H. Zhou, X. Huang, Intrusion detection system combined enhanced random forest with SMOTE algorithm, EURASIP J. Appl. Signal Process. 2022 (1) (2022) 1–20. Dec.
[39] T. Rymarczyk, E. Kozłowski, G. Kłosowski, K. Niderla, Logistic regression for machine learning in process tomography, Sensors 19 (15) (2019) 3400. Aug. 2019.
[40] S. Azam, M. Bibi, R. Riaz, S.S. Rizvi, S.J. Kwon, Collaborative learning based sybil attack detection in vehicular AD-HOC networks (VANETS), Sensors 22 (18) (2022) 6934. Sep. 2022.
[41] S. Orozco-Arias, J. S. Piña, R. Tabares-Soto, L. F. Castillo-Ossa, R. Guyot, and G. Isaza, Measuring Performance Metrics of Machine Learning Algorithms for Detecting and Classifying Transposable Elements.
[42] A. Rácz, D. Bajusz, and K. Héberger, Molecules Multi-Level Comparison of Machine Learning Classifiers and Their Performance Metrics.
[43] E. Paolini, L. Valcarenghi, L. Maggiani, N. Andriolli, Real-time clustering based on deep embeddings for threat detection in 6G networks, IEEE Access 11 (2023) 115827–115835, https://doi.org/10.1109/ACCESS.2023.3325721.
[44] M.S. Rocha, G.D.G. Bernardo, L. Mundim, B.B. Zarpelão, R.S. Miani, Supervised machine learning and detection of unknown attacks: an empirical evaluation, in: L. Barolli (Ed.), Advanced Information Networking and Applications, Lecture Notes in Networks and Systems, vol. 654, Springer, Cham, 2023, https://doi.org/10.1007/978-3-031-28451-9_33. AINA 2023.
[45] D.G. Berbecaru, S. Giannuzzi, D. Canavese, Autoencoder-SAD: an autoencoder-based model for security attacks detection, in: 2023 IEEE Symposium on Computers and Communications (ISCC), 2023, pp. 758–763, https://doi.org/10.1109/ISCC58397.2023.10217930. Gammarth, Tunisia.
[46] J. Zhao, Trust Management Model of VANETs Based on Machine Learning and Active Detection Technology, 2022.
[47] M Sivaram, D Yuvaraj, Machine learning based DDoS detection[J] (2020).