Contents lists available at ScienceDirect

Cyber Security and Applications

journal homepage: http://www.keaipublishing.com/en/journals/cyber-security-and-applications/

Securing the road ahead: Machine learning-driven DDoS attack detection in VANET cloud environments

Himanshu Setia[a], Amit Chhabra[b], Sunil K. Singh[b], Sudhakar Kumar[b,*], Sarita Sharma[c], Varsha Arya[d,k], Brij B. Gupta[e,h,i,j,**], Jinsong Wu[f,g]

[a] Department of CSE, Chandigarh College of Engineering and Technology, Sector 26, Chandigarh, India
[b] Department of CSE, Chandigarh College of Engineering and Technology, Chandigarh, India
[c] Department of ECE, Chandigarh College of Engineering and Technology, Chandigarh, India
[d] Department of Business Administration, Asia University, Taiwan, China
[e] Department of Computer Science and Information Engineering, Asia University, Taichung 413, Taiwan, China
[f] School of Artificial Intelligence, Guilin University of Electronic Technology, 510004, China
[g] Department of Electrical Engineering, University of Chile, 8370451, Chile
[h] Department of Electrical and Computer Engineering, Lebanese American University, Beirut, 1102, Lebanon
[i] Symbiosis Centre for Information Technology (SCIT), Symbiosis International University, Pune, India
[j] School of Computing, Skyline University College, P.O. Box 1797, Sharjah, United Arab Emirates
[k] Center for Interdisciplinary Research, University of Petroleum and Energy Studies (UPES), Dehradun, India

ARTICLE INFO

Keywords:
DDoS
VANET
Cloud
NS2
Machine learning
Fuzzification
Simulation

ABSTRACT

Vehicular ad-hoc network (VANET) technology has gained prominence, especially in the context of the emerging field of VANET Cloud as an integral part of connected and autonomous vehicles. The automotive industry's move towards automation and the integration of vehicles into the digital ecosystem has revolutionized wireless network communications. Nevertheless, security remains a paramount concern in these advanced technological landscapes. Safeguarding system integrity and data privacy is of utmost importance before the widespread adoption of VANET Cloud solutions. This study addresses the critical challenge of security within the context of VANET Cloud. Specifically, the focus is on anticipating and mitigating Distributed Denial of Service (DDoS) attacks, which can potentially disrupt the functioning of connected vehicles and associated cloud-based services. To tackle this issue, an innovative architectural framework is proposed to capture and analyze network flows within the VANET Cloud environment. Additionally, it leverages machine learning techniques for classification and predictive analytics with an accuracy of 99.59%. The architecture presented in this research offers the potential to significantly enhance security measures in VANET Cloud deployments. Its adaptability ensures practical applicability to real-world systems, enabling timely responses to security threats and breaches.

1. Introduction

With an increasing reliance on online technology, there has been a surge in disruptive attempts to compromise vital web technologies, leading to service outages. As it continues to discover new types of attacks that disrupt networks, network security has become a matter of national importance [1]. Due to the ever-changing objectives of attackers and the tools they employ, these attacks are constantly evolving. Even now, individuals responsible for a variety of attacks, often involving botnets, resulting in property damage totaling tens of thousands of dollars, are being apprehended and brought to trial [2,3].

Web servers that operate on the Hypertext Transfer Protocol (HTTP) are vulnerable to DoS attacks. The Denial-of-Service attack is one of the earliest forms of cyber-attacks on the internet, primarily intended to render a device unavailable to legitimate users by depleting computer resources. Presently, Denial of Service remains a major concern for cybersecurity experts [4–7]. Distributed denial-of-service attacks, characterized by their lack of statistical patterns in attack behavior and their reliance on standard protocols and services, have supplanted

Peer review under responsibility of KeAi Communications Co., Ltd.

* Corresponding author.

** Corresponding author at: International Center for AI and Cyber Security Research and Innovations, & Department of Computer Science and Information Engineering, Asia University, Taichung 413, Taiwan.

E-mail addresses: amitchhabra@ccet.ac.in (A. Chhabra), sksingh@ccet.ac.in (S.K. Singh), sudhakar@ccet.ac.in (S. Kumar), saritasharma@ccet.ac.in (S. Sharma), bbgupta@asia.edu.tw (B.B. Gupta), wujs@ieee.org (J. Wu).

https://doi.org/10.1016/j.csa.2024.100037
Received 17 December 2023; Received in revised form 4 January 2024; Accepted 9 January 2024
Available online 17 January 2024

traditional peer-to-peer attack methods [8–10]. DoS attackers inundate target servers with a high volume of unwanted data packets, causing network congestion and exhausting the processing capabilities of the server [11]. A notable example of this is the DDoS attack, where numerous devices simultaneously target a specific user [12–14]. DDoS attacks demand substantial computing resources, resulting in higher costs and making them relatively easier to detect [15,16].

At present, Denial of Service remains a prominent concern for cybersecurity experts, even in the context of Cloud computing [17–19]. Distributed denial-of-service attacks, which lack statistical rules for attack behavior and exploit standard protocols and services, have replaced traditional peer-to-peer attack methods [20]. DoS attackers inundate target servers with a high volume of unwanted data packets, causing network congestion and ultimately exhausting the processing capability of the server. In the realm of VANET Cloud, these challenges take on new dimensions.

The DDoS attack, demanding a significant amount of PC resources, incurs relatively high costs, rendering it easier to detect [21,22]. This research within the VANET Cloud framework aims to address these issues comprehensively. Parallel computing [23–27] can be used to simulate and model traffic scenarios, considering the dynamic interactions among connected vehicles to enhance processing power.

In this paper, the integration of VANETs within the context of VANET Cloud is explored, forming a dynamic network of moving and stationary vehicles connected by a wireless infrastructure [28]. Nodes within this network share information seamlessly, contributing to the collective intelligence of the system. The implementation employs NS2, a discrete event simulation system renowned for its support in emulating and analyzing diverse protocols across both local and satellite networks.

This paper is structured into five sections, aligning with the conventions of scholarly literature. The first section delves into the current state of research on DDoS attack detection, outlining the motivation behind this study and identifying key limitations. The second section discusses approach to simulating DDoS attacks within the VANET Cloud framework, encompassing the methodology, training models, and the application of fuzzification techniques for enhancement. The third section provides a comprehensive description of the calculations involved, including detailed explanations of the formulas employed, parameters utilized, and an overview of the dataset. It also includes an audit of the metrics under scrutiny and the generation of graphical representations across various parameters.

Subsequently, the results section presents the culmination of this research, consolidating findings and their implications within the VANET Cloud framework. Finally, the concluding section summarizes contributions and outlines potential avenues for future research within the dynamic landscape of VANET Cloud security.

2. Literature review

DDoS attacks has been a significant challenge in network security. Traditional detection methods involve gathering traffic characteristics and applying various algorithms to confirm attack presence. When an attack is identified, countermeasures are implemented, often at firewalls or other protective devices, leading to the identification and dropping of attack traffic. However, these methods often rely on specific attack characteristics, making them susceptible to evasion by attackers who can spoof source IP addresses [1,29,30].

In the early 2010s, the less secure internet architecture, characterized by simpler designs and reduced core network complexity, favored attackers. Core routers lacked the capability to authenticate incoming IP packets, resulting in IP spoofing as a primary source of DDoS attacks [4].

Researchers have also explored alternative methods, including Self-Organizing Maps (SOM) and Machine Learning (ML) [31–33] algorithms for DDoS attack prevention. In these approaches, SOM processes flow statistics (with various parameters) gathered from OpenFlow switches. However, these methods often require extensive computation time for large matrix computations in the SOM training model, as well as for ML models [2,15].

Advancements have also been made in the field of cloud computing. In a recent study [34], authors proposed a model for a Selective Cloud Egress Filter (SCEF) that identifies various types of attacks by analyzing relevant packet metadata and takes preventive actions when detecting a DDoS attack on virtual machines (VMs) [35]. The effectiveness of attack detection depends on the characteristics of the specific attack.

However, analyzing VM bundle data alone may not provide a reliable method for identifying ongoing attacks, as communication gaps can lead to spikes in packet rates. Software-defined networking (SDN) [18] has gained widespread acceptance in the network community, with SDN-based network security solutions actively under development. SDN, in combination with techniques such as entropy analysis, machine learning algorithms, and traffic pattern analysis, offers the potential for more efficient DDoS attack detection [5]. However, these SDN-based frameworks are still in their early stages, and their accuracy requires further validation.

Methods like IP traceback and packet filtering have also been studied for DDoS detection, involving packet marking to trace the attack path and filtering techniques to remove malicious traffic. However, the limited length of the IP identification field, capped at 16 bits, may prove insufficient to store the entire path [36].

The most used mechanism for DDoS attack detection involves analyzing traffic attributes before implementing any preventive measures. Presently, DDoS attack prevention relies on network intrusion detection [37]. After sufficient information about the attack is collected, attack traffic is blocked and dropped [38]. Various methods, such as those based on entropy [3,39], activity profiling [16], and attacker address distribution [8,21], rely on attack attributes for detection. However, attackers can manipulate these attributes, posing a challenge.

Additional methods, such as Round-Trip Time (RTT) analysis utilizing spectrum analysis and Fourier Power Spectrum Entropy (FPSE) for LDoS attack detection, have been developed and tested on both simulation environments like NS2 and real network traffic [40,41]. Artificial neural networks combined with wavelet energy spectrum analysis have been applied for classifying network traffic, with tests conducted on FTP servers [42,43]. Unsupervised detection, aided by autoencoders, has exhibited great accuracy in identifying low-rate traffic attacks [44]. The MIPDV metric was introduced to identify DDoS attacks, and simulations on NS2 demonstrated its potential for achieving higher accuracy [45].

Machine learning-based DDoS attack detection has also made progress, employing algorithms such as the Naive Bayesian algorithm, hidden Markov model, and support vector machine [46]. Despite advancements in DDoS attack defense at all levels of the SDN architecture, significant challenges remain in accurately identifying DDoS attacks [47,48].

In this study [49], the authors addressed the critical issue of road safety, particularly in scenarios with low visibility. They introduced an innovative algorithm that modifies the AODV routing protocol within VANETs Cloud. This algorithm alerts drivers when a leading vehicle in the same lane decelerates, potentially revolutionizing accident prevention in challenging road conditions. In this study [50], researchers focused on enhancing security and efficiency within the VANET-Cloud environment. They proposed the DEMD2RS VANET-Cloud' framework. Key components of their approach included the introduction of the Hash-based Credential Authentication Scheme (HCAS) to ensure security for both Road Side Units (RSUs) and vehicles. Additionally, the study employed the Firm Aware Clustering Scheme (FACS) using the Stud Krill Herd (SKH) algorithm to maintain cluster stability. Data retrieval was facilitated through Twofish encryption, and an Artificial Neural Network (ANN) algorithm was employed for efficient path determination. To reduce delay in emergency message dissemination, the authors utilized the Fuzzy-Topsis (FT) algorithm for selecting the best disseminator.

Table 1
Related Works.

Ref.	Year	Approach	Advantages	Disadvantages
34	2023	DEMD2RS VANET-Cloud framework includes HCAS, FACS, Twofish, ANN, and FT processes.	Enhanced security through SHA-3 and ECP authentication in HCAS.	Potential computational overhead due to the cryptographic operations involved in HCAS, which may impact system performance.
33	2023	Modifies the AODV routing protocol within the VANETs Cloud to provide cautionary notifications to vehicles when a front-running car in the identical lane decelerates.	Enhanced road safety by alerting drivers to decelerating vehicles, aiding better decisions and accident prevention in low visibility.	Infrastructure and compatibility issues with existing vehicular communication systems.
32	2021	Naive Bayesian algorithm, hidden Markov model, support vector machine used to detect DDoS attacks	Detect DDoS attacks	Significant shortcomings in identifying DDoS attacks
7	2020	Identifies various attack types by examining packet metadata, and takes preventative measures when a DDoS attack is discovered on virtual machines.	A DDoS attack has been identified on virtual machines (VMs).	Detection of attack subject to characteristics of attack, analysis of VM bundle data may not be reliable for detecting attacks.
8	2017	Use techniques such as entropy, machine learning algorithms, traffic pattern analysis to detect DDoS attacks	The examination for detecting DDoS attacks.	Early stage of development, accuracy not yet confirmed
13	2013	Detect DDoS attacks by analysing traffic attributes, prevent attacks by collecting information and blocking/dropping attack traffic	Prevent attacks by collecting information and blocking/dropping attack traffic	Attackers can easily fool methods based on attack attributes
6	2010	ML algorithms for DDoS attack prevention	DDoS attack prevention	Hours of computation required for SOM training and machine learning models

The considerations and surveys in the field of DDoS attack detection and prevention provide valuable insights [51]. This research aims to contribute to this domain by training a model through simulations conducted on NS2. Simulating two different scenarios using the software and collecting their information in a file would enable us to utilize relevant features for classifying or predicting a DDoS attack in an unseen scenario (Table 1).

3. DDOS attack trends

At the very onset of DDoS attacks, a significant form of attack involved the generation of an extensive volume of traffic with maximum throughput. In the year 2000, websites like Amazon and Yahoo were targeted by this DDoS attack. For this form of attack, a variety of attack tools, including TFN, Trinoo, TFN2K, and Stacheldracht, were used. Detection of this attack relied on network traffic anomaly detection techniques. While it was possible to detect such attacks, accurately distinguishing attack packets proved to be extremely challenging, mainly due to the IP address spoofing techniques employed by attackers.

In the mid-2000s, another variant of DDoS attacks emerged. During this period, hackers sought to recruit as many zombie PCs as possible, as a larger number of compromised systems translated into a higher volume of attack traffic. For this purpose, attackers developed Internet Worms capable of autonomously and rapidly gathering zombie PCs. This led to a widespread Internet Worm pandemic, targeting vulnerable systems, and subsequently generating an extensive DDoS attack, exemplified by the Slammer Worm. The Slammer Worm spread rapidly across the world and caused significant disruption.

A significant DDoS assault that targeted 48 sites in the USA and South Korea between July 5 and July 10, 2009, used numerous zombie computers to attempt a variety of explicit attacks, including TCP SYN flooding, UDP/ICMP flooding, HTTP GET flooding, and CC attacks. Protecting HTTP services from assaults while these attacks were being tried proved useless. This is explained by the fact that nearly every DDoS identification approach depends on the variety and quantity of data being broadcast. The volume of attack traffic from each assault system was insufficient, therefore even if the total quantity of attack traffic was massive, these approaches were unable to identify the distinct attackers.

4. DDOS attack and proposed methodology

4.1. DDOS attack in VANET cloud

VANET Cloud can be visualized as the network of interconnected vehicles communicating asynchronously and independently as shown in Fig. 1. The three primary components of VANET Cloud are On-Board Unit (OBU), Application Unit (AU), and Road-Side Unit (RSU). OBUs and AUs are found on the network nodes that each vehicle is represented as, much like PC networks. Since these networks are designed to be self-

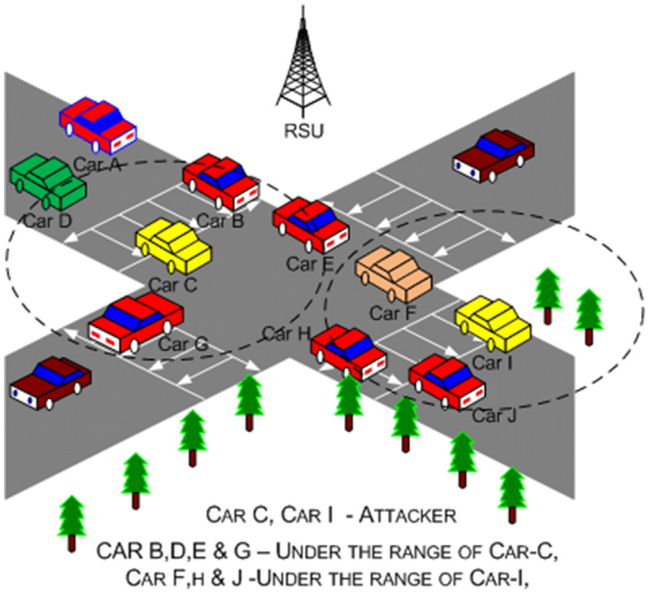

RSU
Car A
Car B
Car D
Car C
Car E
Car F
Car G
Car H
Car I
Car J
Car C, Car I - Attacker
CAR B,D,E & G – Under the range of Car-C,
Car F,H & J -Under the range of Car-I,

Fig. 1. Real life VANET Cloud visualization. [Source 2: Upadhyaya, Ajay. (2018). ATTACKS ON VANET SECURITY].

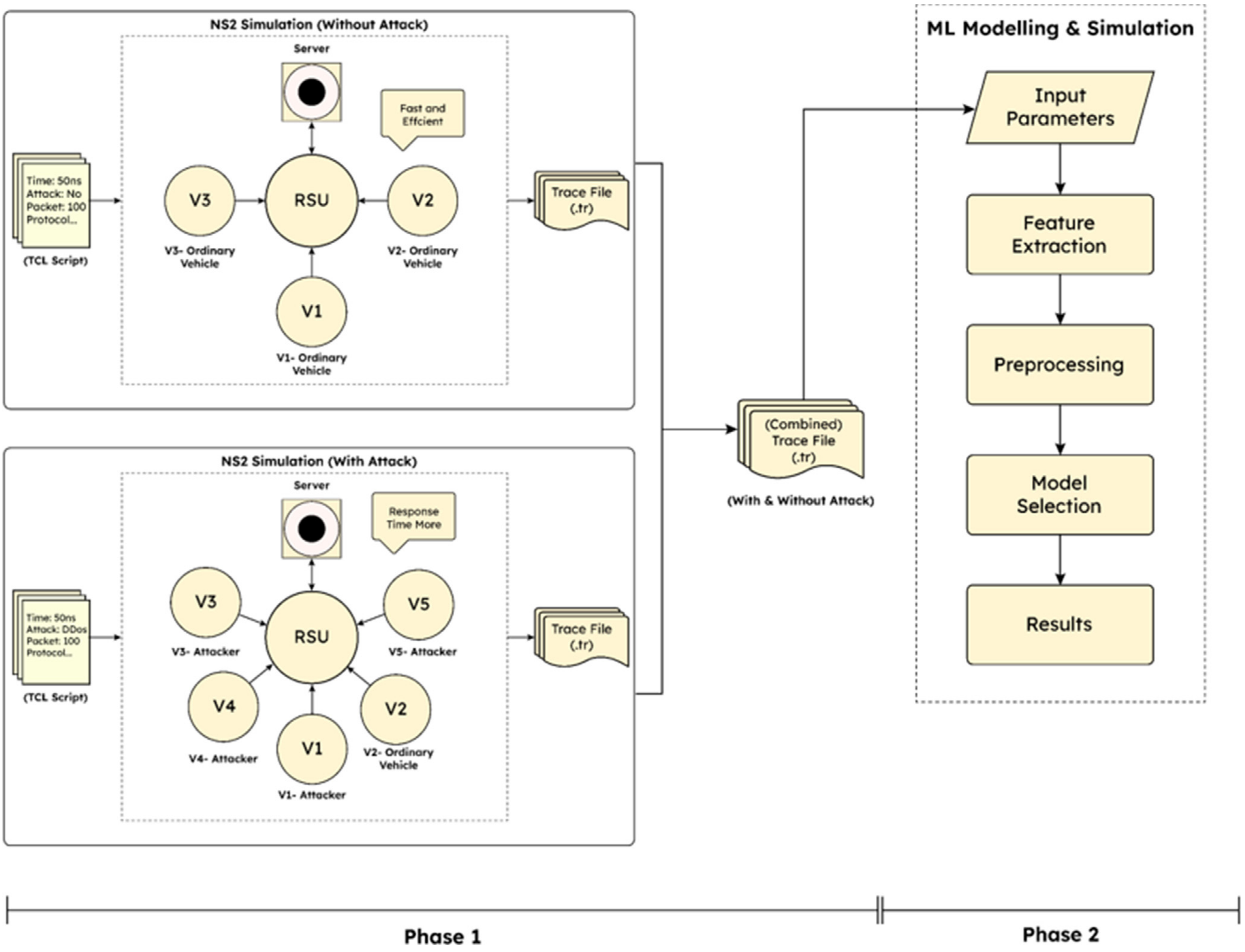


Fig. 2. Proposed Methodology.

organizing and provide ad hoc services, they are known as "ad-hoc" networks.

Due to the periodicity of transmissions and short contention window sizes, VANET Cloud is susceptible to synchronization-based DDoS assaults. Vehicles can communicate with one another within their range because to its architecture. If a vehicle is out of communication range, data is delivered via a multi-hop approach to the appropriate destination node. Road-side unit communication enables cars to send, receive, and relay data to other vehicles by expanding the range of network connectivity. The network nodes in VANET Cloud are decentralized, and their topologies are constantly changing, making it challenging to identify malicious attacks, broken nodes, and compromised vehicles.

4.2. Proposed methodology

In the subsequent Subsections, a structured methodology is proposed, consisting of two distinct phases, both contributing significantly to the overarching objective of enhancing security and safety within vehicular networks. The approach considers the dynamic and challenging nature of vehicular environments, with a particular emphasis on real-time communication and proactive measures for accident prevention.

4.2.1. Phase 1- NS2 simulation

Two NS2 simulations were conducted: one without any attack, involving only ordinary nodes engaged in normal communication, and another with an attack scenario where attacker nodes were introduced to disrupt communication and divert packets, affecting the packet delivery ratio.

Initially, a TCL script was written, specifying parameters such as communication type, MAC protocol, routing protocol, etc., as detailed in the paper. Both scenarios, with and without an attack, were simulated for a duration of 50 ns, with a maximum of 100 packets. An AWK script was employed to calculate the Packet Delivery Ratio (PDR) in both simulations, and the results were recorded in the Trace file.

In the scenario without an attack, ordinary vehicles (represented as nodes) communicated efficiently with the Road-Side Unit (RSU), which was connected to the main server. Communication flow was smooth, resulting in a significant PDR value. Conversely, in the attack scenario, attacker nodes disrupted the normal packet flow by diverting them, creating a DDoS situation. Communication in this case experienced delays, leading to a reduced PDR.

Throughout the simulation, various parameters including Source IP, Destination IP, time frame, packet length, and PDR were timestamped and stored in a Trace (.tr) file. The trace files from both scenarios were merged to generate a dataset, completing the first phase of this process as shown in 2 - PHASE-1.

The following parameters shown in Table 2 were utilized in the TCL script to simulate the DDoS attack on VANET Cloud, along with their respective values:

Table 2
Simulation Parameters.

Parameter	Value
Channel Type	Wireless Channel
Radio Propagation Model	Two Ray Ground
Network Interface Type	WirelessPhy
MAC Type	802_11
Interface Queue Type	DropTail/ PriQueue
Link Layer Type	LL
Antenna Model	OmniAntenna
Max Packets	100
Number of mobile nodes	4
Routing Protocol	AODV
X dimension of topography	500
Y dimension of topography	500
Time of Simulation	50ns

4.2.2. *Phase 2- prognostication using machine learning models*

After merging both the trace files and converting it into.csv format, a dataset has been generated. It is now ready to be used for applying different types of classification models such as KNN, Random Forest, Decision Tree etc. Using the input parameters (as mentioned in the 'Dataset Description' part below), classification will be done whether the scenario belongs to "Normal" or "DDoS."

Prior to applying machine learning models, data preprocessing steps, as depicted in Fig. 3, are necessary to handle incomplete, redundant, and inconsistent data. In this step, filtering redundant data and normalizing the dataset should be conducted. For feature extraction, the focus will be on identifying and utilizing the most relevant attributes from the dataset. This process aims to eliminate irrelevant and redundant information before employing classifiers. Ultimately, following completion of all data preprocessing steps, various classification models will be implemented on the generated dataset, and accuracy evaluation will be carried out, as indicated in Fig. 2 (PHASE-2).

The whole process is shown in the form of flow chart in Fig. 4. For comparison purposes, the confusion matrices have been plotted and performance metrics such as Accuracy score, Precision score etc. have been calculated.

The two main classes of supervised machine learning algorithms used for prediction are regression and classification algorithms. While classification methods anticipate the output categorical values, regression approaches predict the output continuous values. The primary goal of this study is to forecast the categorical values of DDoS and benign attacks on target labels in the dataset. Among the most popular classification algorithms are Logistic regression, Random Forest, Decision Tree, Naive Bayes, K-Nearest Neighbour, and AdaBoost. These techniques are not only faster but also substantially more accurate than traditional techniques for detecting a DDoS attack.

4.3. *Fuzzification*

Concept of fuzzification is used for categorizing the degree of correlation for cases involving values between upper and lower attributes bound. For preserving proposed model accuracy and better mechanism fuzzified algorithm had been constructed and applied (refer to Algorithm 1).

Simple model of fuzzification is used as shown Fig. 5. As, the value obtained cannot be deduced to elite extent and to remove the errors the concept of fuzzy set is used. A fuzzy set can be defined as a charting of a set of real numbers (xi) onto membership values (ui) which commonly fall in the span of [0, 1] and it is useful tool to portray circumstances in which the data is unclear. It levers by assigning a degree to which a specific object fit in set. Suppose value obtained is 0.424 then by using fuzzy set the descriptive value can be obtained like – 0.424 belongs to low fuzzy set which is having a membership function of 0.1 and it also

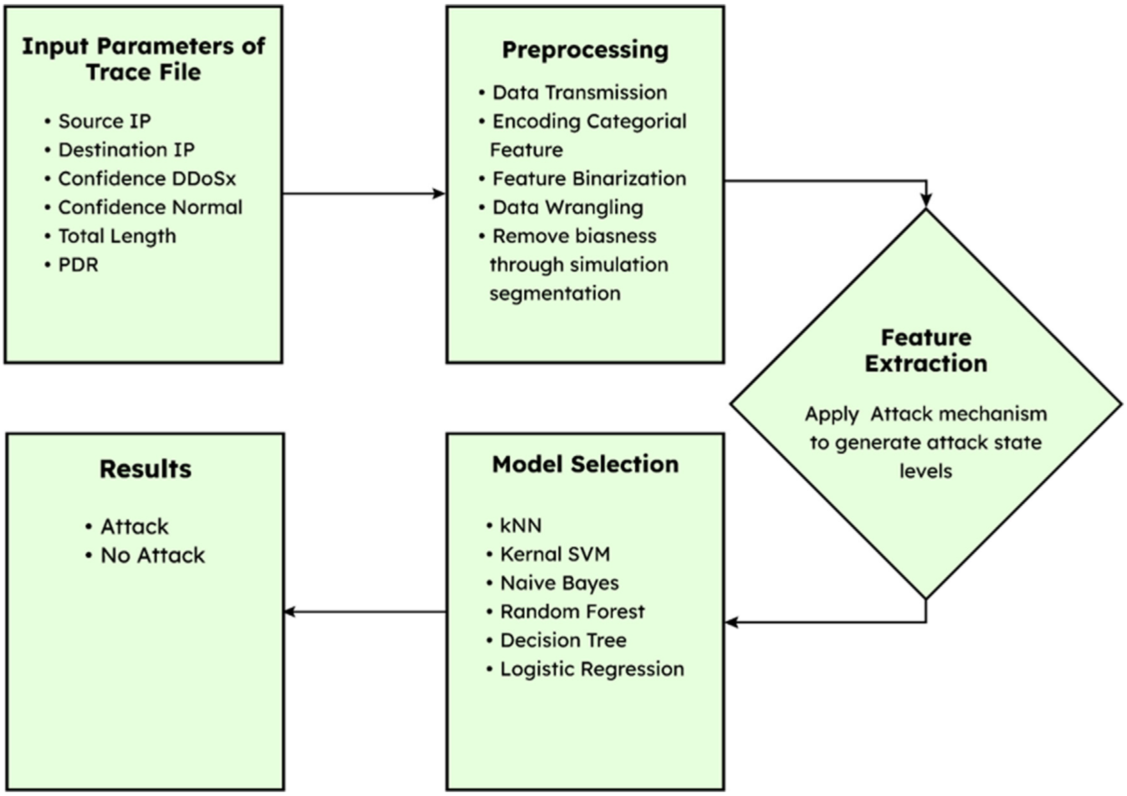


Fig. 3. Workflow of machine learning.

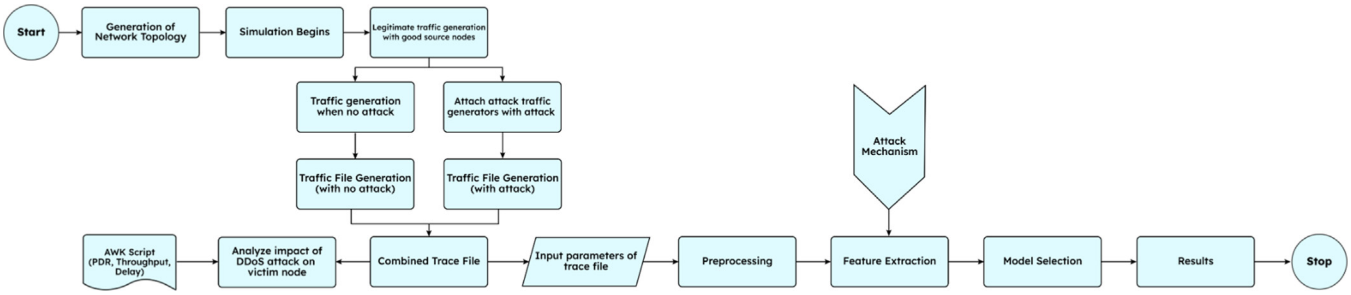

Start
Generation of Network Topology
Simulation Begins
Legitimate traffic generation with good source nodes
Traffic generation when no attack
Attach attack traffic generators with attack
Traffic File Generation (with no attack)
Traffic File Generation (with attack)
Attack Mechanism
AWK Script (PDR, Throughput, Delay)
Analyze impact of DDoS attack on victim node
Combined Trace File
Input parameters of trace file
Preprocessing
Feature Extraction
Model Selection
Results
Stop

Fig. 4. Methodology flowchart.

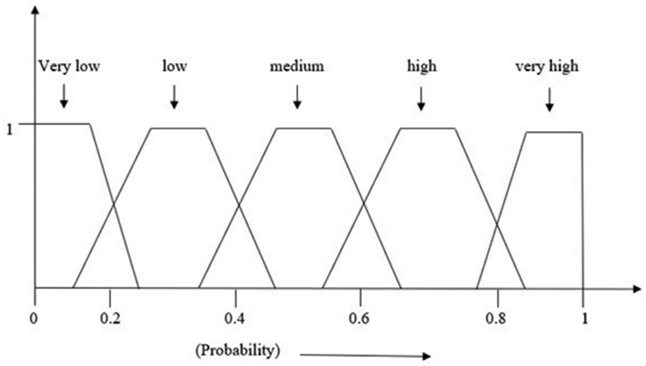


Fig. 5. Fuzzification model.

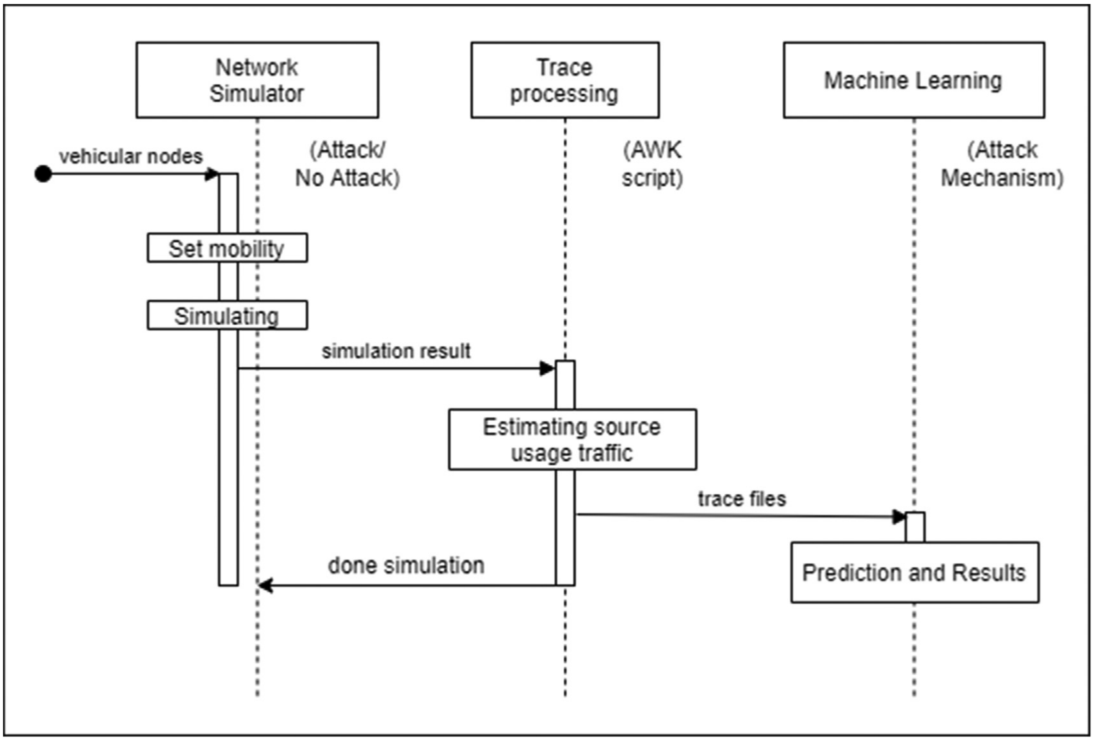


Fig. 6. Sequence diagram of implementationl.

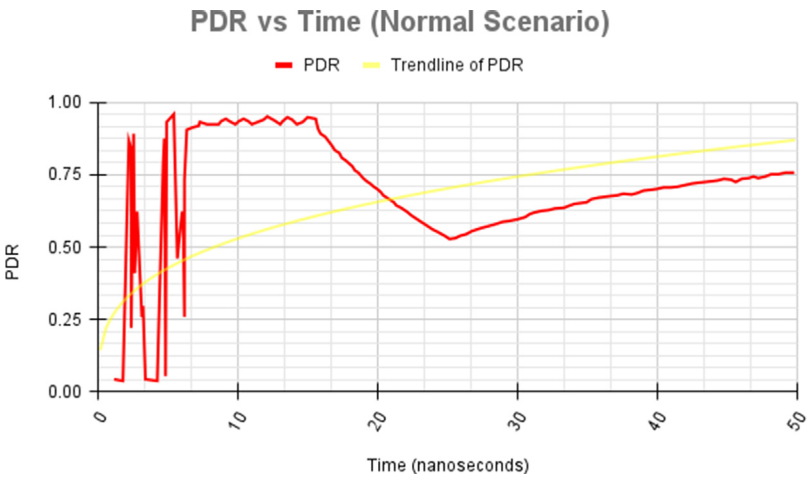


Fig. 7. PDR vs Time Graph (Normal Scenario).

belongs to medium fuzzy set with a membership function of 0.9. Thus, on basis of priority the précised area i.e., can be fetched.

Algorithm 1: Fuzzified DDoS detection algorithm

```
Input: Set of tuples of VANET parameters T = (A1, S1, D1, CD1, CN1, L1, P1), (A2, S2, D2, CD2, CN2, L2, P2), (A3, S3, D3, CD3, CN3, L3, P3), …… (An, Sn, Dn, CDn, CNn, Ln, Pn)
Where A1 is the attack type, S1 is the source IP, D1 is the destination IP, CD¬1 is the Confidence of DDoS scenario, Cn¬1 is the Confidence of Normal scenario, L1 is the total length and P1 is the packet delivery ratio.
Output: Set of processed tuples T = (A1, CD1, CN1, L1, P1), (A2, CD2, CN2, L2, P2), (A3, CD3, CN3, L3, P3), …… (An, CDn, CNn, Ln, Pn) and result of classification as “Attack” or “Normal”
Where A1 is the attack type, CD¬1 is the Confidence of DDoS scenario, Cn¬1 is the Confidence of Normal scenario, L1 is the total length and P1 is the packet delivery ratio.
function classifyAttack(T)
S = tokenize T to T1, T2, T3, ….. Tn
for i = 1 to N do
    remove non dependent parameters from tokens T1,T2,T3,….Tn store the dependent parameters for implementing models; Split the tokens into training set and test set; Train the model using the training set and test its accuracy;
end
Input a tuple T;
Apply classification model and display output;
```

5. Calculations and result

All experiments have been conducted on the Linux platform [33]. Certain calculations have been conducted by precisely following the sophisticated sequence of steps performed during implementation as shown in Fig. 6.

The trendline graphs are generated from the simulation conducted in NS2. Based on which certain results are concluded.

The graph shown in Fig. 7 depicts the plotting between packet delivery ratio and time in case of normal scenario. The trendline of PDR follows a smooth curvature. The actual PDR values during the initial timeframe of 10 nanoseconds varies abruptly between 0 and 1 due to initial communication between nodes but thereafter follows a path like the power series trendline.

The graph shown in Fig. 8 plotted of PDR vs Time in attack scenario illustrates the proposed model performance. The trendline of PDR in such cases is usually a semi-curved straight line. In proposed VANET model, the PDR values are initially less but then gradually the plot goes above the trendline, giving better results. The PDR values increases after a certain time interval as the nodes can classify the attack scenario and counter the malicious nodes.

The graph shown in Fig. 9 depicts the energy values with respect to time in a normal scenario is somewhat like the earlier one of PDR vs Time. When the PDR values were changing rapidly, the energy values also show the similar behavior. After a certain time, value, both the graphs showed a smooth curvature.

In graph shown in Fig. 10, the energy values according to the trendline will be very high at the starting because of high attack intensity but then start decreasing following a negative exponential curve. In proposed model, the energy values remain almost constant till a certain timestamp and then after attack classification happens, the values are better as compared to the trendline. Further is description about attributes of dataset (refer to Table 3).

The attributes having major impact are PDR and energy. They can be derived as-

$$PDR = \frac{\text{Number of packets received}}{\text{Number of packets transmitted}}$$

$$Energy = \frac{\text{Unit energy per packet *no. of packets received*hop count}}{\text{PDR}}$$

The dataset describes the various attributes on which model has been trained and are shown in Fig. 11. Some of main points drawn from dataset are mean value of PDR in combined dataset is 0.845 (approx.). Despite this there is underlying pattern in packet length in combined dataset that is 75.399 (approx.). Same for the case in energy usage which is 108.399 (approx.) for combined dataset.

Some of traditional score calculation methods and terms are used for evaluation purposes of different models which are-

Fig. 12 depicts the confusion matric for various models and Table 4 shows the accuracy, precision, recall and F1 score obtained for various models. The Decision Tree and Random Forest models achieve the highest accuracy, precision, recall, and F1 scores, while the Kernel SVM model demonstrates the lowest scores. Logistic Regression, Naive Bayes, and k-NN models exhibit performance levels between those of the Decision Tree/Random Forest models and the Kernel SVM model. Both the Decision Tree and Random Forest models provide the highest accuracy, making them suitable for predicting simulation-based attacks.

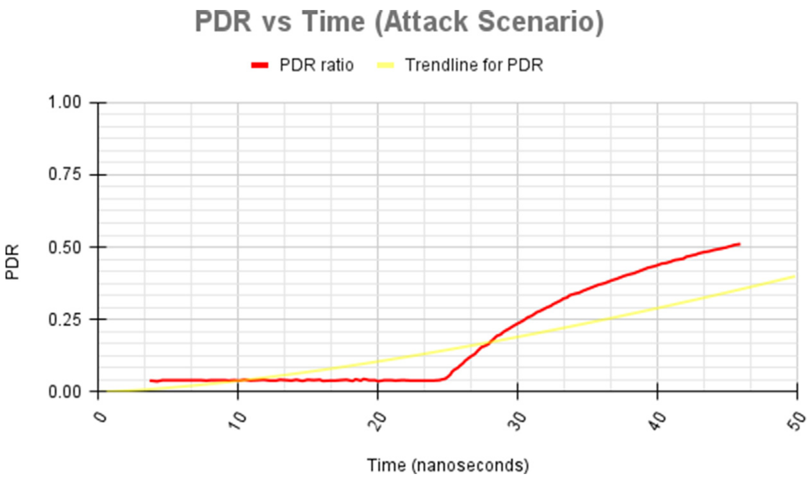


Fig. 8. PDR vs Time Graph (Attack Scenario).

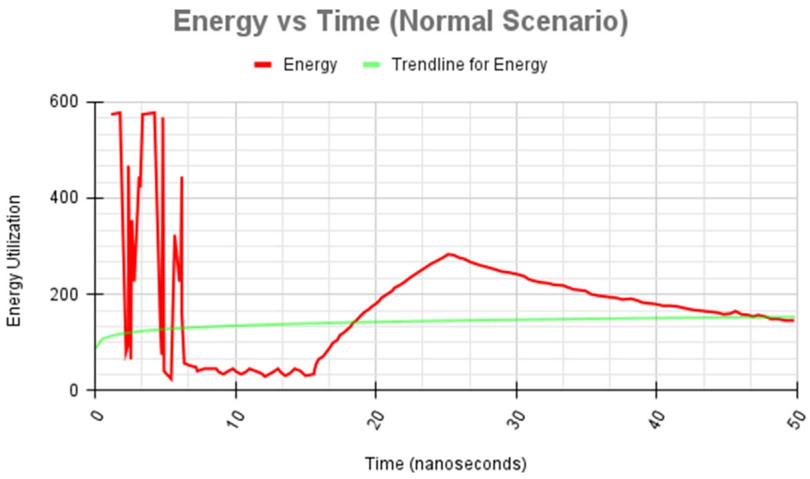


Fig. 9. Energy vs Time Graph (Normal Scenario).

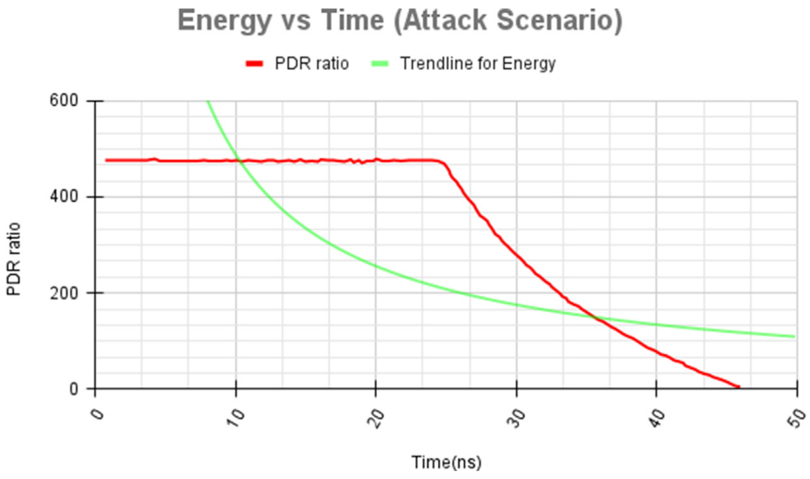
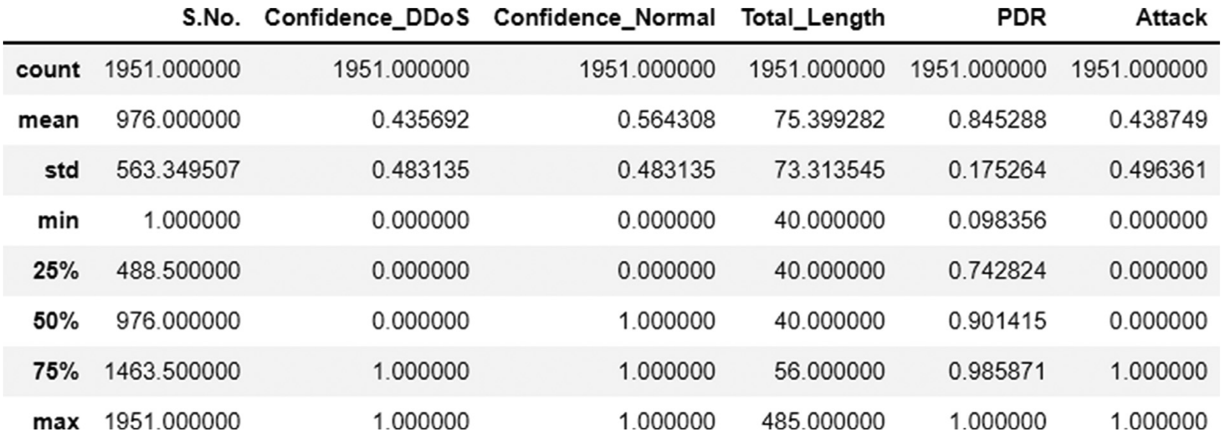


Fig. 10. Energy vs Time Graph (Attack Scenario).

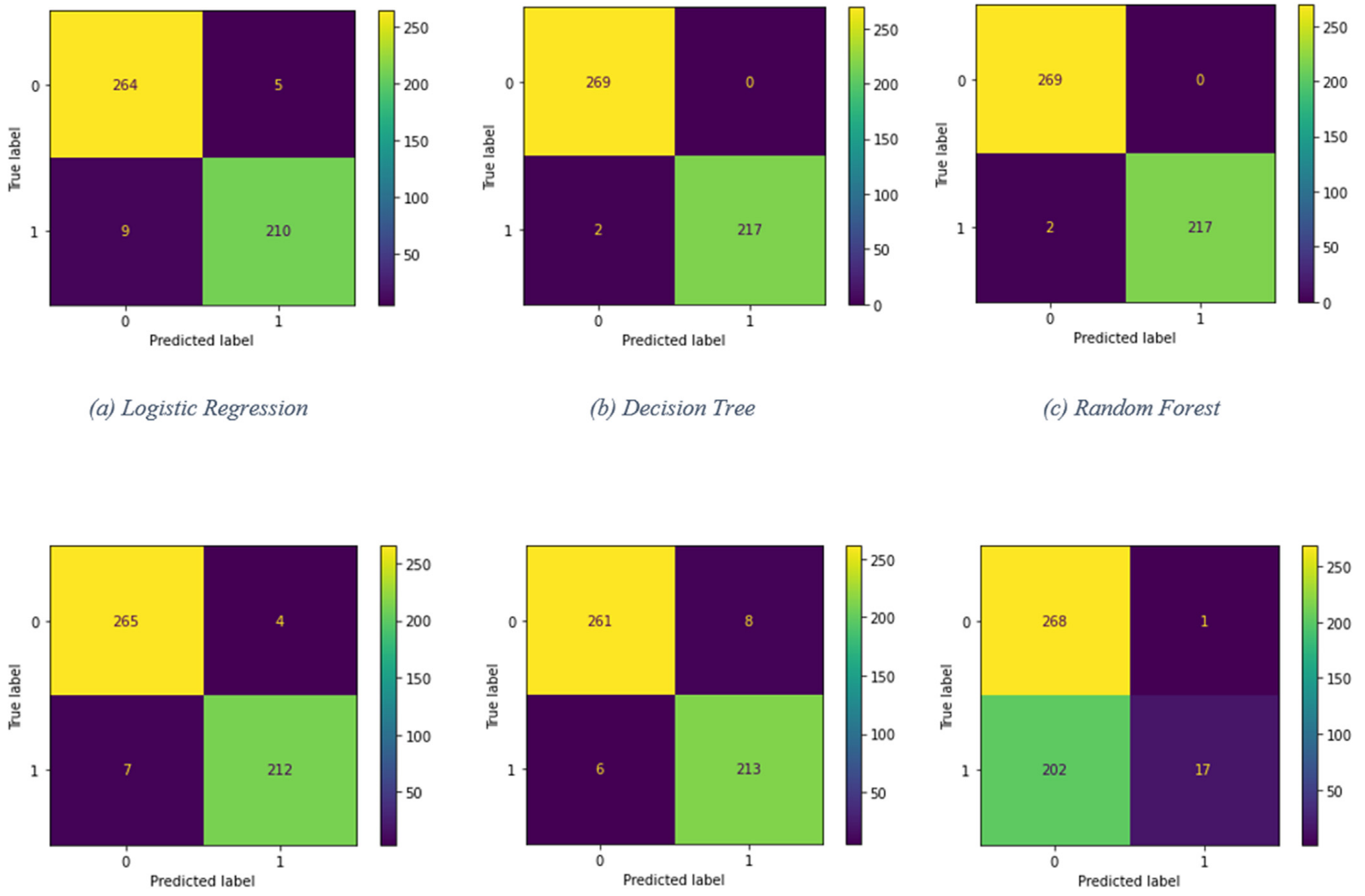
Table 3
Dataset Attributes List.

Attribute	Description
Attack	The data collected is from a normal scenario or from a DDoS scenario
Source IP	IP address of the source location
Destination IP	IP address of the destination location
Confidence DDoS	The confidence value of the scenario being a DDoS attack
Confidence Normal	The confidence value of the scenario being a normal VANET simulation
Total Length	Length of packets being sent
Packet Delivery Ratio	$\frac{\text{Packets sent successfully}}{\text{Total packets sent}}$



	S.No.	Confidence_DDoS	Confidence_Normal	Total_Length	PDR	Attack
count	1951.000000	1951.000000	1951.000000	1951.000000	1951.000000	1951.000000
mean	976.000000	0.435692	0.564308	75.399282	0.845288	0.438749
std	563.349507	0.483135	0.483135	73.313545	0.175264	0.496361
min	1.000000	0.000000	0.000000	40.000000	0.098356	0.000000
25%	488.500000	0.000000	0.000000	40.000000	0.742824	0.000000
50%	976.000000	0.000000	1.000000	40.000000	0.901415	0.000000
75%	1463.500000	1.000000	1.000000	56.000000	0.985871	1.000000
max	1951.000000	1.000000	1.000000	485.000000	1.000000	1.000000

Fig. 11. Dataset parameters distribution.



(a) Logistic Regression

(b) Decision Tree

(c) Random Forest

(d) kNN Classification

(e) Naïve Bayes

(f) Kernel SVM

Fig. 12. Confusion matric of various classification models.

Table 4
Performance metrics table.

Classification Model	Accuracy Score	Recall Score	Precision Score	F1 Score
Logistic Regression	0.9713	0.9701	0.9718	0.9709
Decision Tree	0.9959	0.9954	0.9963	0.9958
Random Forest	0.9959	0.9954	0.9963	0.9958
kNN	0.9774	0.9765	0.9778	0.9771
Naive Bayes	0.9713	0.9714	0.9706	0.9710
Kernel SVM	0.5840	0.5369	0.7573	0.4343

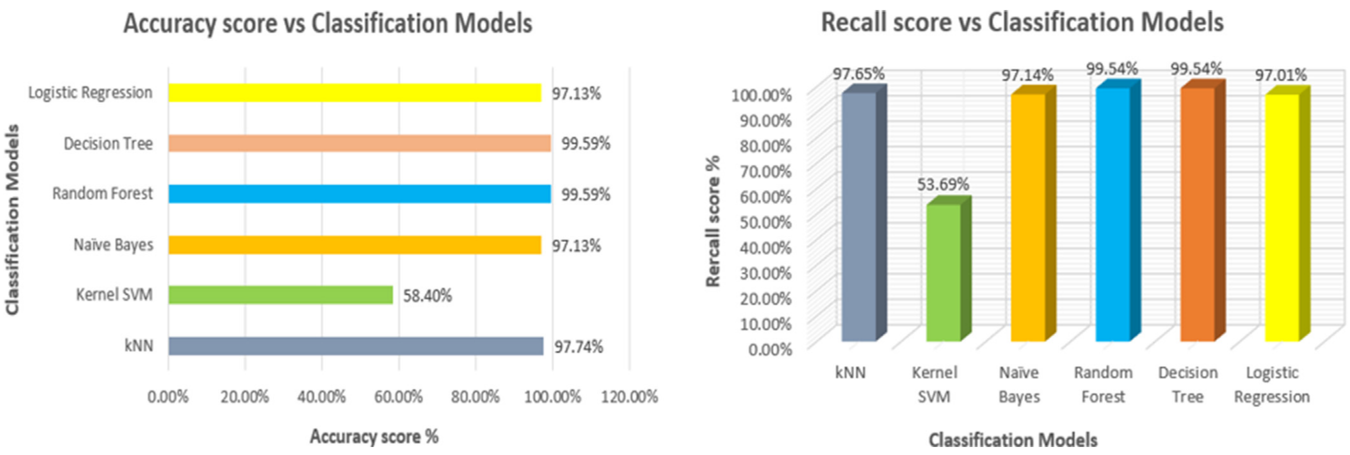


Fig. 13. Accuracy vs Classification Models and Recall score vs Classification Models.

The log-odds of an event are expressed as a linear combination of one or more independent variables in the logistic model, a statistical approach used to calculate the likelihood of an event occurring. In the case of logistic regression, the true negative value is 264, and the true positive value is 210. There are 5 false positives and 9 false negatives.

Decision trees and random forests exhibit the highest accuracy, with a true negative value of 269 and a true positive value of 217. There are no false positives and 2 false negatives.

K-Nearest Neighbors (KNN) is a proximity-based classification method used to predict the arrangement of individual data points. In this dataset, the true negative value is 265, and the true positive value is 212. There are 4 false positives and 7 false negatives.

Naive Bayes is a probabilistic classifier that makes predictions based on the likelihood of an item's existence. In this context, it has a true negative value of 261 and a true positive value of 213, with 8 false positives and 6 false negatives.

In case of Kernal SVM, the true negative value is 268, the true positive value is 17, with 1 false positive and 202 false negatives. It is important to note that the performance in terms of true positives for this model is significantly lower compared to other models.

One parameter (refer to Table 4 and Fig. 13) for assessing classification models is accuracy. Notably, among the machine learning models evaluated, the random forest and decision tree both have the best accuracy scores of 0.9959 of all the models mentioned above [31–33]. The accuracy score of the KNN is 0.9774 at this moment, while the accuracy scores of logistic regression and naive Bayes classifier are both 0.9713. The accuracy score for Kernel SVM is the lowest, at 0.5840.

The recall score is a metric for determining how accurately a model counts actual positive values among all real positive values. The highest recall score of 0.9954 is shared by the decision tree and random forest, as seen in Table 4. Recall scores for Logistic Regression, KNN, and Naive Bayes were 0.9714, 0.9765, and 0.9701 respectively [52,53]. The recall score for Kernel SVM is the lowest (0.5369).

In terms of positive observations, precision is defined as the ratio of accurately predicted observations to all expected positive observations. The precision scores for decision trees and random forests are both 0.9963. Precision scores for Logistic Regression, KNN, and Naive Bayes were 0.9718, 0.9706, and 0.9706 respectively. The precision score of 0.7573 for Kernel SVM is the lowest.

The F1-score measures a model's accuracy on a dataset [54]. As the harmonic mean of recall and precision, it is so named. Similarly, the decision tree and random forest both have the greatest F1 scores of 0.9958, while other models have respectable overall F1 scores with Kernel SVM scoring the lowest (0.4343).

6. Conclusion

This study presents a pioneering approach to the detection of DDoS attacks within the context of Vehicular Ad Hoc Networks (VANET Cloud). Notably, this research stands as a singular exploration of ML based methodologies for DDoS attack detection within VANET Cloud environments. A substantial contribution of this investigation is the extensive statistical analysis of network traffic characteristics, encompassing both normative network states and adversarial attack scenarios. This rigorous analysis has facilitated the establishment of essential criteria and thresholds for the accurate identification of potential DDoS attack risks. The simulation results divulge a significant inverse correlation between energy consumption and packet delivery rate (PDR) within the VANET Cloud framework. Moreover, the Random Forest and Decision Tree machine learning models have demonstrated noteworthy predictive capabilities, yielding elevated detection rates. Furthermore, the future scope of the research involves creating synthetic datasets through Generative Adversarial Networks (GANs) to elevate the accuracy of machine learning models in DDoS attack detection. This approach aims to create more realistic and diverse datasets for enhanced model preparedness against a wide array of potential threats. Subsequent research will focus on a meticulous comparison of these models via systematic hyperparameter tuning, optimizing their performance and adaptability. In summation, this research advances the domain of DDoS attack detection within VANET Clouds by harnessing the power of machine learning. The discernments derived from this study contribute to the ongoing efforts to bolster the security and resilience of VANET Clouds against the backdrop of evolving cyber threats.

Declaration of competing interest

Brij B. Gupta is an associate editor for Cyber Security and Applications and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

CRediT authorship contribution statement

Himanshu Setia: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Amit Chhabra:** Supervision, Project administration. **Sunil K. Singh:** Writing – review & editing, Supervision, Project administration, Investigation, Data curation, Conceptualization. **Sudhakar Kumar:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sarita Sharma:** Supervision, Conceptualization. **Varsha Arya:** Supervision, Conceptualization. **Brij B. Gupta:** Supervision, Project administration, Investigation, Conceptualization. **Jinsong Wu:** Investigation, Funding acquisition, Conceptualization.

Acknowledgement

This research work is supported by National Science and Technology Council (NSTC), Taiwan Grant No. NSTC112- 2221-E-468-008-MY3.

References

[1] Z. Duan, X. Yuan, J. Chandrashekar, Controlling IP spoofing through interdomain packet filters, IEEE Trans. Dependable Secure Comput. 5 (2008) 22–36.

[2] S. Ostermann, B. Tjaden, M. Ramadas, Detecting anomalous network traffic with self-organizing maps, in: Recent Advances in Intrusion Prevention, 2003, pp. 36–54.

[3] M. Poongodi, V. Vijayakumar, F. Al-Turjman, M. Hamdi, M. Ma, Intrusion prevention system for DDoS attack on VANET with reCAPTCHA controller using information based metrics, IEEE Access 7 (2019) 158481–158491, doi:10.1109/ACCESS.2019.2945682.

[4] T. Peng, C. Leckie, K. Ramamohanarao, Survey of network-based defense mechanisms countering the dos and DDoS problems, ACM Comput. Surv. 39 (1) (2007) 3.

[5] N.Z. Bawany, J.A. Shamsi, K. Salah, DDoS attack detection and mitigation using SDN: methods, practices, and solutions, Arab. J. Sci. Eng. 42 (2017) 425–441.

[6] A.K. Goyal, G. Agarwal, A.K. Tripathi, Network architectures, challenges, security attacks, research domains and research methodologies in VANET: a survey, Int. J. Comput. Netw. Inf. Secur. 11 (10) (2019) 37–44, doi:10.5815/ijcnis.2019.10.05.

[7] M. Singh, S.K. Singh, S. Kumar, U. Madan, T. Maan, Sustainable framework for metaverse security and privacy: opportunities and challenges, in: In International Conference on Cyber Security, Privacy and Networking, Cham: Springer International Publishing, 2021, pp. 329–340.

[8] A. Sinha, P.S.K. Mishra, Preventing VANET from DOS & DDOS attack, Int. J. Eng. Trends Technol. (IJETT) V4 (10) (2013) 4373–4376. ISSN:2231-5381. www.ijettjournal.org. published by seventh sense research group

[9] Z. Lu, G. Qu, Z. Liu, A survey on recent advances in vehicular network security, trust, and privacy, IEEE Trans. Intell. Transp. Syst. 20 (2) (2019), doi:10.1109/TITS.2018.2818888.

[10] B. Zhang, T. Zhang, Z. Yu, DDoS detection and prevention based on artificial intelligence techniques, in: Proc. 3rd IEEE International Conference on Computer and Communications (ICCC), 2017, pp. 1276–1280.

[11] X. Wang, Z. Ning, M.C. Zhou, X. Hu, L. Wang, Y. Zhang, F.R. Yu, B. Hu, Privacy-preserving content dissemination for vehicular social net-works: challenges and solutions, IEEE Commun. Surveys Tuts. 21 (2) (2018) 1314–1345. 2nd Quart.

[12] X. Yuan, C. Li, X. Li, DeepDefense: identifying DDoS attack viadeep learning, in: Proc. IEEE International Conference on Smart Computing, 2017, pp. 1–8.

[13] S.X. Wu, W. Banzhaf, The use of computational intelligence in intrusion detection systems: a review, Appl. Soft Comput. 10 (1) (2010) 1–35.

[14] K. Aggarwal, S.K. Singh, M. Chopra, S. Kumar, F. Colace, Deep learning in robotics for strengthening industry 4.0.: opportunities, challenges and future directions, in: Robotics and AI for Cybersecurity and Critical Infrastructure in Smart Cities, 2022, pp. 1–19.

[15] E. Mota, A. Passito, R. Braga, Lightweight DDoS flooding attack prevention using NOX/openflow, in: IEEE 35th conference on Local Computer Networks, 2010, pp. 408–415.

[16] K. Adhikary, S. Bhushan, S. Kumar, K. Dutta, Evaluating the impact of DDoS attacks in vehicular ad-hoc networks, Int. J. Secur. Privacy Pervasive Comput. 12 (2020) 1–18, doi:10.4018/IJSPPC.2020100101.

[17] T. Saini, S. Kumar, T. Vats, M. Singh, Edge computing in cloud computing environment: opportunities and challenges, in: International Conference on Smart Systems and Advanced Computing (Syscom-2021), 2020.

[18] R. Singh, S.K. Singh, S. Kumar, S.S. Gill, SDN-aided edge computing-enabled AI for IoT and smart cities, in: SDN-Supported Edge-Cloud Interplay for Next Generation Internet of Things, 2022, pp. 41–70.

[19] F.J.G. Pećalvo, A. Sharma, A. Chhabra, S.K. Singh, S. Kumar, V. Arya, A. Gaurav, Mobile cloud computing and sustainable development: Opportunities, challenges, and future directions, Int. J. Cloud Appl. Comput. (IJCAC) 12 (1) (2022) 1–20.

[20] L. Liu, Y. Wang, J. Zhang, Q. Yang, A secure and efficient group key agreement scheme for VANET, Sensors (Switzerland) 19 (3) (2019), doi:10.3390/s19030482.

[21] H. Karthikeyan, G. Usha, Real-time DDoS flooding attack detection in intelligent transportation systems, Comput. Electr. Eng. 101 (2022) 107995, doi:10.1016/j.compeleceng.2022.107995. ISSN 0045-7906

[22] X. Hu, J. Zhao, B.C. Seet, V.C.M. Leung, T.H.S. Chu, H. Chan, S-Aframe: agent-based multilayer framework with context-aware semantic service for vehicular social networks, IEEE Trans. Emerg. Top. Comput. 3 (1) (2015) 44–63.

[23] S. Kumar, S.K. Singh, N. Aggarwal, K. Aggarwal, Evaluation of automatic parallelization algorithms to minimize speculative parallelism overheads: an experiment, J. Discrete Math. Sci. Cryptogr. 24 (5) (2021) 1517–1528.

[24] S. Kumar, S.K. Singh, N. Aggarwal, B.B. Gupta, W. Alhalabi, S.S. Band, An efficient hardware supported and parallelization architecture for intelligent systems to overcome speculative overheads, Int. J. Intell. Syst. 37 (12) (2022) 11764–11790.

[25] S.S. Kumar, S.K. Singh, N. Aggarwal, K. Aggarwal, Efficient speculative parallelization architecture for overcoming speculation overheads, in: International Conference on Smart Systems and Advanced Computing (SysCom-2021), 3080, 2021, pp. 132–138.

[26] S. Kumar, S.K. Singh, N. Aggarwal, Speculative parallelism on multicore chip architecture strengthen green computing concept: a survey, in: Advanced Computer Science Applications, Apple Academic Press, 2023, pp. 3–16.

[27] S.K. Singh, Linux Yourself: Concept and Programming, CRC Press, 2021.

[28] X. Hu, J. Deng, J. Zhao, W. Hu, E.C.-H. Ngai, R. Wang, J. Shen, M. Liang, X. Li, V.C.M. Leung, Y.K. Kwok, SAfeDJ: a crowd-cloud codesign approach to situation-aware music delivery for drivers, ACM Trans. Multimed. Comput. Commun. Appl. 12 (1s) (2015) 21, doi:10.1145/2808201.

[29] X. Ma, Y. Chen, DDoS prevention method based on chaos analysis of network traffic entropy, IEEE Commun. Lett. 18 (1) (2014) 114–117.

[30] D. Moore, C. Shannon, D.J. Brown, G.M. Voelker, S. Savage, Inferring internet denial-of-service activity, ACM Trans. Comput. Syst. 24 (2) (2006) 115–139.

[31] I. Singh, S.K. Singh, R. Singh, S. Kumar, Efficient loop unrolling factor prediction algorithm using machine learning models, in: 2022 3rd International Conference for Emerging Technology (INCET), IEEE, 2022, pp. 1–8.

[32] F.J.G. Pećalvo, T. Maan, S.K. Singh, S. Kumar, V. Arya, K.T. Chui, G.P. Singh, Sustainable stock market prediction framework using machine learning models, Int. J. Softw. Sci. Comput. Intell. (IJSSCI) 14 (1) (2022) 1–15.

[33] G. Mengi, S.K. Singh, S. Kumar, D. Mahto, A. Sharma, Automated machine learning (autoML): the future of computational intelligence, in: International Conference on Cyber Security, Privacy and Networking, Cham: Springer International Publishing, 2021, pp. 309–317 .

[34] G.I. Shidaganti, A.S. Inamdar, S.V. Rai, A.M. Rajeev, SCEF: a model for prevention of DDoS attacks from the cloud, Int. J. Cloud Appl. Comput. (IJCAC) 10 (3) (2020) 67–80, doi:10.4018/IJCAC.2020070104.

[35] A. Singh, B.B. Gupta, Distributed denial-of-service (DDoS) attacks and defense mechanisms in various web-enabled computing platforms: issues, challenges, and future research directions, Int. J. Semant. Web Inf. Syst. (IJSWIS) 18 (1) (2022) 1–43.

[36] A. Yaar, A. Perrig, D. Song, Pi: a path identification mechanism to defend against DDoS attack, in: Proceedings of the IEEE Symposium on Security and Privacy, 2003, pp. 93–107.

[37] Z. Ling, Z.J. Hao, Intrusion detection using normalized mutual information feature selection and parallel quantum genetic algorithm, Int. J. Semant. Web Inf. Syst. (IJSWIS) 18 (1) (2022) 1–24.

[38] S. Parkinson, P. Ward, K. Wilson, J. Miller, Cyber threats facing autonomous and connected vehicles: future challenges, IEEE Trans. Intell. Transp. Syst. 18 (11) (2017) 2898–2915.

[39] Y. Gao, H. Wu, B. Song, Y. Jin, X. Luo, X. Zeng, A distributed network intrusion detection system for DDoS detection in VANET, IEEE Access 7 (2019), doi:10.1109/ACCESS.2019.2948382. 1-1

[40] H. Kaur, K. Saluja, S. Behal, Analysis of web services under HTTP attack using real time testbed, Int. J. Control Theory Appl. 9 (2016) 279–292.

[41] Z. Chen, C.K. Yeo, B.S. Lee, C.T. Lau, Power spectrum entropy based detection and mitigation of low-rate dos attacks, Comput. Netw. 136 (2018) 80–94, doi:10.1016/j.comnet.2018.02.029. ISSN 1389-1286

[42] M. Yue, L. Liu, Z. Wu, M. Wang, Identifying LDoS attack traffic based on wavelet energy spectrum and combined neural network, Int. J. Commun. Syst. 31 (2017), doi:10.1002/dac.3449.

[43] T. Zhao, D.C.-T. Lo, K. Qian, A neural-network based DDoS detection system using hadoop and HBase, in: IEEE International Conference on High Performance Computing and Communications, 2015, pp. 1326–1331.

[44] B.A. Pratomo, P. Burnap, G. Theodorakopoulos, Unsupervised approach for detecting low rate attacks on network traffic with autoencoder, in: 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), 2018, pp. 1–8, doi:10.1109/CyberSecPODS.2018.8560678.

[45] M. Şimşek, A new metric for flow-level filtering of low-rate DDoS attacks, Secur. Commun. Netw. 8 (2015), doi:10.1002/sec.1302.

[46] G. Ajeetha, G. Priya, Machine learning based DDoS attack detection, 2019, 1-5. 10.1109/i-PACT44901.2019.8959961.

[47] L.F. Eliyan, R.D. Pietro, Dos and DDoS attacks in software defined networks: a survey of existing solutions and research challenges, Future Gener. Comput. Syst. 122 (2021) 149–171, doi:10.1016/j.future.2021.03.011. ISSN 0167-739X

[48] A. Mishra, B.K. Joshi, V. Arya, A.K. Gupta, K.T. Chui, Detection of distributed denial of service (DDoS) attacks using computational intelligence and majority vote-based ensemble approach, Int. J. Softw. Sci. Comput. Intell. (IJSSCI) 14 (1) (2022) 1–10.

[49] K.T. Chui, T.S. Kochhar, A. Chhabra, S.K. Singh, D. Singh, D. Peraković, V. Arya, Traffic accident prevention in low visibility conditions using VANETs cloud environment, Int. J. Cloud Appl. Comput. (IJCAC) 12 (1) (2022) 1–21.

[50] M. Al-Shabi, An efficient delay aware emergency message dissemination and data retrieval in secure VANET-cloud environment, Wirel. Pers. Commun. (2023) 1–36.

[51] A. Gaurav, B.B. Gupta, P.K. Panigrahi, A novel approach for DDoS attacks detection in COVID-19 scenario for small entrepreneurs, Technol. Forecast. Soc. Change 177 (2022) 121554.

[52] I. Singh, S.K. Singh, S. Kumar, K. Aggarwal, Dropout-VGG based convolutional neural network for traffic sign categorization, in: Congress on Intelligent Systems: Proceedings of CIS 2021, vol. 1, Singapore: Springer Nature Singapore, 2022, pp. 247–261.

[53] T. Vats, S.K. Singh, S. Kumar, et al., Explainable context-aware IoT framework using human digital twin for healthcare, Multimed. Tools Appl. (2023) 1–25.

[54] S. Gupta, S. Agrawal, S.K. Singh, S. Kumar, A novel transfer learning-based model for ultrasound breast cancer image classification, in: Computational Vision and Bio-Inspired Computing: Proceedings of ICCVBIC 2022, Singapore: Springer Nature Singapore, 2023, pp. 511–523.