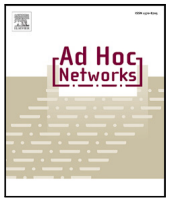
Ad Hoc Networks 140 (2023) 103026

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc



RBF-SVM kernel-based model for detecting DDoS attacks in SDN integrated vehicular network

Goodness Oluchi Anyanwu [a], Cosmas Ifeanyi Nwakanma [b], Jae-Min Lee [a], Dong-Seong Kim [a,*]

[a] Networked System Laboratory, Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
[b] ICT Convergence Research Center, Kumoh National Institute of Technology, Gumi, South Korea

ARTICLE INFO

Keywords:
SDN-based VANET
SVM kernels
GridSearch CV
Hyper-parameter optimization
Intrusion detection system

ABSTRACT

The development of the intelligent transport space (ITS) comes with the challenge of securing transportation data. As the vehicular network is highly dynamic, the network architecture is susceptible to distributed malicious attacks irrespective of the emergence and integration of enabling technologies. A Software Defined Network (SDN) based VANET is an improvement over the traditional VANET and may be susceptible to attacks as a result of its centralized structure, resulting in dangerous circumstances. SDN-based VANET security for 5G deployments is essential and calls for integrating artificial intelligence tools for threat detection. An intrusion detection system was suggested in this work to detect Distributed Denial of Service (DDoS) attacks in the vehicular space. The proposed solution uses the Grid Search Cross-Validation (GSCV) exhaustive parameter search technique and the Radial Basis Function Kernel of the Support Vector Machine (RBF-SVM) algorithm. The performance of other ML algorithms implemented was compared using key performance metrics. Experimental simulations to validate the proposed framework's efficacy against DDoS intrusion show that the proposed scheme demonstrated an overall accuracy of 99.40% and a mean absolute error of 0.006.

1. Introduction

With diverse worldwide components and cutting-edge technologies, it is possible to access the VANET via a wide range of devices [1]. VANET has been enhanced by various sophisticated technologies, including edge computing, cloud computing, and Software-Defined Networking (SDN) [2]. The SDN controls the whole network methodically, outperforming the conventional VANET and serving as an enabler for the construction of 5G networks. It offers VANET unique scaling and centralized control. Nevertheless, the deployment and integration of new entities and architectural components into any system introduces security risks and vulnerabilities [3]. As a result, several types of malicious interference can affect SDN-based VANETs. Consequently, VANET becomes an easy target for attackers considering the integrated SDN architecture's single point of control.

Traditionally, attacks are designed to disrupt high-speed networks, flooding the principal control medium with large volume of malicious data [4]. In SDN-based VANETs, attackers seek to disable the SDN architecture by focusing on the SDN controller, resulting in unavailable services and degraded performance, culminating in a Distributed Denial of Service (DDoS) [5]. A hacked controller can cause a DDoS attack. A high frequency of illicit connection requests originates from many sources, confounding and causing issues for legitimate users. DDoS is a significant sort of assault in VANET. It has a significant influence on a network's availability. The most common DDoS assaults occur at several levels, taking advantage of flaws in network connectivity, preventing information exchange, and rendering the network unresponsive [6].

DDoS attacks are a threat to SDN-based VANET infrastructures. A DDoS attack on the SDN controller simultaneously targets its process and communication capabilities. As a result, protecting VANETs from DDoS assaults is a challenging task [7]. Because these networks function in real-time, the use of encryption and authentication for intrusion detection systems (IDSs) is unreliable. These types of IDS solely focus on the data's payload, lowering the size of the purported data. However, machine learning (ML) models have been used to identify and categorize many sorts of dispersed hostile intent on VANET and associated technologies, as these approaches are less influenced by encryption, securing the vehicular space. Furthermore, these techniques strive for high dependability and in-depth analysis of the input data [8].

In this work, for reliable DDoS detection accuracy and a robust learning mechanism, the Radial Basis Function (RBF) kernel of SVM is assisted by an exhaustive Grid Search Cross-Validation (GSCV) for

* Corresponding author.
E-mail addresses: anyanwu.goodnes@kumoh.ac.kr (G.O. Anyanwu), cosmas.ifeanyi@kumoh.ac.kr (C.I. Nwakanma), ljmpaul@kumoh.ac.kr (J.-M. Lee), dskim@kumoh.ac.kr (D.-S. Kim).

https://doi.org/10.1016/j.adhoc.2022.103026
Received 12 March 2022; Received in revised form 18 October 2022; Accepted 23 October 2022
Available online 7 November 2022

optimizing SVM parameters. Different kernel functions have dissimilar learning and generalization abilities. Although the RBF kernel is associated with its suitability for non-linear data, it has no basis for the selection and combination of optimal parameter functions towards effective performance. To assist the RBF kernel, the GSCV implements an exhaustive exploration over a defined kernel parameter grid to reach cross-validated and optimized parameter values [9]. Moreover, the advantage of the RBF kernel over other kernels is its efficient utilization of computing memory [10]. The experimental results obtained using the publicly available DDoS evaluation dataset for SDN architectures from IEEE dataport [11] show that the proposed RBF-SVM outperformed other state-of-the-art ML algorithms, the default SVM, SVM kernel variants, and other parameter selections in terms of classification accuracy as well as with better generalization.

The proposed model can be embedded within an SDN-based VANET to define security rules for preventing DDoS attacks. It proves to be an effective solution for protecting VANET from these attack types. The novel contributions of this work are as follows:

1. Proposing a novel extension of the regular RBF-SVM kernel-based detection approach. The proposed RBF-SVM kernel structure is based on parameter tuning of the basic RBF-SVM kernel to reach an optimal selection towards enhancing the detection rate and accuracy of identifying DDoS attacks.
2. Critical and extensive review of recent IDS based on ML and DL approaches. The majority of the papers reviewed are selected based on similar proposed ML methods (SVM) and comparison evaluation with a focus on DDoS detection, datasets used, optimization approach employed, and target system as VANET or SDN-VANET.
3. To enhance the kernel parameters of the RBF-SVM kernel, an exhaustive GSCV optimization technique is employed over five-fold cross-validation (CV). In this way, hyper-parameters with the best performance for the RBF-SVM classifier were obtained, and the most accurate score was determined for the DDoS attack classification.
4. The performance of the proposed methods was then evaluated using the SDN-DDoS dataset after an in-depth analysis of the dataset properties. For verification of the experimental approach and to access the performance of the proposed model under scaling conditions, the CICDDoS dataset was further used for evaluation.
5. Ultimately, a comprehensive and comparative investigation into other SVM kernel variants, default and random parameter selections of the RBF-SVM kernel and selected state-of-the-art ML algorithms are also provided in this work for better performance analysis.

The remainder of the paper is organized as follows: Section 2 examines current VANET DDoS attack detection and mitigation options, whereas Section 3 presents our suggested research methodology and specifics of our implementation process. Section 4 describes the assessment metrics. Section 5 presents the collected findings and the effectiveness of our proposed approach over existing methods in detecting and mitigating DDoS assaults. Finally, Section 6 brings the article to a close.

2. Related works

A vast volume of research has emphasized the importance of secured vehicular communication, presenting IDSs that use either ML or Deep Learning (DL) methodologies. Random Forest (RF), k-Nearest Neighbor (kNN), and Support Vector Machine (SVM) are the most implemented supervised ML approaches employed as these methods guarantee reliable outcomes [12]. However, SVM remains one of the ubiquitous approaches that has found relevance in detecting intrusion for VANET, achieving reliable performances [13].

Authors in [14] evaluated selected ML approaches (kNN, Decision Tree (DT), Artificial Neural Network (ANN), and SVM) to detect DDoS attacks in an SDN-enabled network. They employed Neighborhood Component Analysis (NCS) for feature analysis. In addition, they adopted Stochastic Gradient Descent (SGD) to optimize parameters for a better detection process. The dataset used for evaluation is the SDN-DDoS dataset. Their experimental results show that DT had better performance than the other algorithms, having an accuracy rate of 100%. However, DT algorithms are prone to over-fitting. Hence, the reliability of DT performances is always in doubt.

In a similar vein, [15] examined several ML classifiers to identify the best model. A network emulator was used to gather their self-generated dataset. Three phases make up the proposed detection mechanism: flow collection, feature aggregation, and classification. According to experimental findings, the Gradient Boosting Classifier (GBC), among others like Logistic Regression (LR), Naive Bayes (NB), Neural Network (NN), and RF, provided the best results. A Deep Neural Network (DNN) based on a four-layered Stacked Sparse AutoEncoder (SSAE) and softmax classifier was proposed in [16] to identify DDoS assaults in SDN-based VANETs. The SSAE was used to dimension-reduce and recover the most crucial SDN-based VANET features. In addition to SVM, kNN, and DT, other layer variations were also included in the suggested model, which had a 96.9% accuracy rate.

RF was utilized as the proposed classifier to characterize DDoS attacks in [17]. The evaluation was also carried out on their generated dataset using the RBF-SVM algorithm. However, the SVM model produced a high false positive rate (FPR) compared with other techniques. A limitation of the simulation scenario is the small amount of attack traffic generated. A novel systematic approach based on DDoS data was introduced in [18] using the Bayesian Optimization Method (BOM) for determining the optimal hyper-parameter of the DT model. The combination of BOM-DT and Minimum Redundancy Maximum Relevance (MRMR) feature selection produced a 99.35% accuracy score. However, the authors did not compare their idea with existing ML methods that adopted MRMR feature selection for a fair comparison.

Attack detection using a novel combination of SDN and Growing Hierarchical Self-Organizing Maps (GHSOM) is proposed in [19]. The GHSOM analyzes the multilayered HD data by finding structures in it and arranging them in a 2-D output space to address drawbacks encountered during the modeling process. The Network Control Plane (NCP) aligns the intended outcome's dependability and scalability with the demands of VANET-based SDN networks. However, the simulation employed the KDD-CUP 1999 dataset based on a DARPA study for evaluation. The KDD CUP 1999 dataset is frequently used in studies on general IDS studies and may not be well suited for a VANET scenario, limiting the potential of the proposed technique.

The multi-layered SVM in [20] is based on Kernel Principal Component Analysis (KPCA) and Genetic Algorithm (GA). The KPCA was used to reduce feature vector dimensions, while the GA was adopted to optimize various SVM parameters. The improved kernel function reduced the training time of the IDS, achieving an accuracy of 98.07%. The authors of [21] suggested a trigger module based on incoming packets. The proposed module considered the Entropy-based Feature Extraction (EFE) technique to extract features from the SDN network architecture. Based on the kernel's statistical learning theory, various Kernel variants of the SVM model were explored. The classification performance of the linear kernel function using the self-generated dataset provided a 97.68% accuracy in comparison with the RBF at 95.4%. However, optimizing the RBF kernel could have produced a better result.

The hybrid SVM algorithm (Analysis of Variance (ANOVA) and RBF kernel Dot methods) modeled by the authors in [22] yielded appreciable results compared to the single SVM model. Similarly, the authors in [23] implemented an ML-DL model for mitigating DDoS threats on VANET. The hybrid model, a combination of DT and neural network,

Table 1
Comparison of related literature.

Refs.	Year	ML Meths.	Dataset used	Dedicated for SDN network?	Parameter optimization algorithm	Feature analysis technique selection	Acc. using SVM (%)	Comp. Anal.	Future scope?
[14]	2021	kNN, DT, ANN, SVM	SDN-DDoS	Yes	**SGD**	**NCS**	98.59	No	Yes
[15]	2018	GBC, SVM, NN, LR, NB kNN, DT, RF	Self-generated	Yes	Without Parameter Optimization	Without Feature Selection	~94	No	Yes
[16]	2021	DNN, SVM, kNN, DT	Self-generated	Yes	Without Parameter Optimization	**SSAE**	93.07	No	Yes
[17]	2021	J48, SVM, RF, kNN, ANN, NB	Self-generated	Yes	Without Parameter Optimization	**CFS**	97.3	Yes	Yes
[18]	2022	kNN, DT, SVM	Self-generated	Yes	**BOM**	**MRMR**	96.11	Yes	No
[19]	2021	GHSOM	KDD-IDS	No	Without Parameter Optimization	Without Feature Selection	No	No	No
[20]	2020	SVM	NSL-KDD	No	**GA**	**KPCA**	98.07	Yes	Yes
[21]	2018	SVM	Self-generated	Yes	Without Parameter Optimization	**EFE**	95.4	Yes	No
[22]	2020	SVM	Self-generated	No	**AnovaDot & RBFDot**	Without Feature Selection	~98	No	Yes
[23]	2020	DT & NN	Self-generated	No	Without Parameter Optimization	Without Feature Selection	96.40	No	No
[24]	2020	SVM, J48 MLP, RF	CIC DDoS	No	**Random Grid Search**	Without Feature Selection	93.1	No	No
[25]	2019	RBF-SVM	NSL-KDD UNSW-NB15	No	Without Parameter Optimization	Without Feature Selection	86.63 94.07	No	Yes
[26]	2022	DL	NSL KDD BOT-IoT	No	FASMO	Without Feature Selection	93.16 93.20	Yes	Yes
[27]	2019	D2H-IDS	NSL KDD	No	Without Parameter Optimization	ID3	99.43	Yes	No
This paper	**2022**	**RBF-SVM**	**SDN-DDoS**	**Yes**	**GSCV**	**PCA**	**99.40**	**Yes**	**Yes**

achieved an accuracy of 96.40% compared to a single implementation of the different algorithms at 41.77% (DT) and 76.81% (neural network). Despite the high accuracy of the hybrid model, making it a possible consideration for a near-reliable requirement for VANET, it can result in increased computational complexity. In addition, there is room for improvement in the accuracy achieved by the above works as VANET represents a mission-critical system.

In [24], the focus was on low-rate DDoS attacks in a flexible SDN architecture. A low-rate DDoS attack triggers a timeout on specific protocol mechanisms. The authors performed their evaluation on an Open Network Operating System controller running on a Mininet virtual machine to ensure the simulated environment represents a real-world scenario. Using combined optimization algorithms, an accuracy of 93.1% was achieved. Also, RBF-SVM was employed to characterize DDoS attacks on the NSL-KDD and UNSW-NB15 datasets, achieving accuracies of 86.63% and 94.07% in [25]. The following are contributing factors to how well an IDS performs against DDoS attacks.: system architecture, the detection algorithm, and the data characteristics.

Two VANET uncharacterized datasets (BoT-IoT and NSL-KDD) were used to evaluate a hybrid, optimization-driven trust-based secure model. The model was based on a DL model with varying feature sizes [26]. The optimization technique employed is the Fractional Aquila Spider Monkey Optimization (FASMO) technique, an algorithm used for dimensionality reduction. The attack detection mechanism used by the authors in [27] is also a hybrid-based detection solution that uses deep belief and DT-based hybrid IDS (D2H-IDS) for data dimensionality reduction. An Iterative Dichotomiser 3 (ID3) was used to construct the DT for accurate attack classification in a smart city vehicle network. The NSL-KDD dataset was also used by these authors for attack evaluation irrespective of its unreliability. An accuracy of 99.43% was achieved. The unsuitability of the above datasets for VANET representation is the major drawback in these studies.

A Public Key Infrastructure-based signature model (PKI) for secure data communication in the SDN controller was designed for VANET in [28]. The framework is based on a three-way handshake mechanism for secure session establishment and data transfer in the SDN controller. A formal security model was then used to effectively validate the proposed idea based on the SDN-VANET's core hierarchic security features. An Entropy-based Feature Selection (EFS) method was used for traffic classification by measuring the degree of packet randomness in [29]. In addition to evaluating the level of uncertainty in incoming packets, entropy is measured within a predetermined window size. To classify the traffic as normal or attack traffic, the result is then compared to a predetermined threshold.

Conclusively, from the above works, the design of a comprehensive, reliable, and well-analyzed security framework is required for the

successful and secure implementation of VANET and its associated technologies. Hence, improving the reliability of its information exchange is of immense importance, as this network is very dynamic and must respond in real-time [30]. From the research work outlined above, the pervasiveness of SVM is also not in doubt. SVM possesses a solid theoretical basis with evidence of performing incomparably well in various practical and real-world learning tasks. It works by implementing kernel tricks that map data points into new spaces and establish appropriate decision boundaries [31]. This work leverages the reliability and proven effectiveness of the SVM algorithm for classification and detection purposes.

SVM deals with high-dimensional data while providing reliable accuracy [32]. However, despite this accurate classifier for attack detection in vehicular communications, the effect of diverse SVM kernels, selection of hyper-parameters, and data preprocessing decisions on the performance of IDS is an open question for SDN-based VANET. In addition, the relevance of choosing the most suitable kernel and performance parameters of the implemented algorithms with the Optimal Parameter Value (OPV) is a burning issue in research. Thus, in conjunction with the GSCV, this work utilizes the SVM technique as the prime classifier and the RBF kernel option for investigating and classifying DDoS attacks and benign services in SDN-based VANET.

Table 1 provides an extensive review and comparison of the recent and similar ML and DL IDS-based approaches. The papers reviewed in this section are selected based on similar proposed ML methods (SVM) with a focus on DDoS detection, datasets used, optimization approach employed, and target system as SDN-VANET. To the best of our knowledge, no previous literature contributions have attempted mitigating DDoS-based cyber-attacks for SDN VANET with a combination of GSCV and RBF-SVM kernel. Most of the studies are deficient in optimizing their proposed models towards achieving reliable performance for the attack detection tasks.

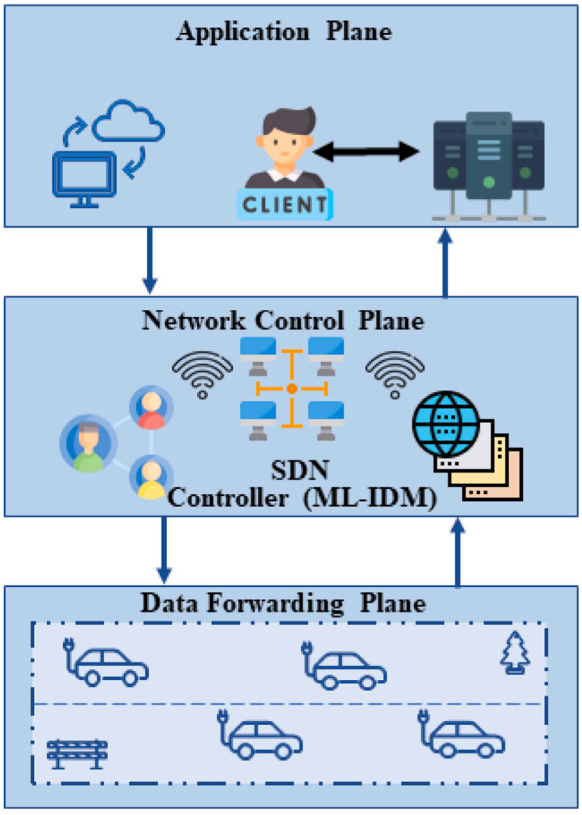


Fig. 1. SDN-based VANET architecture showing proposed DDoS detection scheme.

3. Materials and methods

3.1. Overview of SDN-VANET architecture and attack vulnerabilities

VANET is a self-configuring network that emerged from the advancement of intelligent transport system technology. An accurate and speedy evaluation of any circumstance on the road is one of the benefits VANET offers. The most critical and challenging part of this provision, in addition to the growth of the transportation system, is ensuring it is secure. On the other hand, the SDN framework is an enabling proposal to enhance VANET's general functions. The OpenFlow rules convention utilized by the controllers and incorporated to govern and control VANET supports SDN. Three planes—the application, control, and data planes—make up an SDN-based VANET architecture. Vulnerabilities in any of these planes can be exploited [33].

With the introduction of SDN, independent deployment of processing entities, control, and traffic forwarding are introduced [34]. However, more security complications are introduced as third-party applications liable for security vulnerabilities are introduced. Attackers may overwhelm the controller with traffic from multiple machines, draining its resources. The following are the description of the components of SDN-VANET as shown in Fig. 1.

- The application plane provides services for end users via application programming interfaces. Deployment of various application programs that allow communication and collaboration with all other programs is actualized on this plane.
- SDN-based networks are heavily dependent on the NCP and are responsible for routing and communication. It also provides multiple channels and frequencies. It is made up of modules that coordinate the heterogeneous networks of the intelligent infrastructure, ensuring the distribution of the rules and VANET policy behaviors throughout the network [34].
- The SDN controller is an application in the NCP that provides network traffic flow control for better network administration and application performance. Protocols are used by the SDN controller installed on the NCP to instruct switches on how and where to deliver packets.
- The data forwarding plane is the network architecture layer that physically handles the traffic based on different configurations. It consists of system devices, some of which are routers, switches, and passageways, which are in charge of moving all the client data that flows through the network [6].

3.2. SDN controller-based IDS

A traditional VANET network may comprise an IDS device, often located in a specific area. The IDS limits its detection capabilities to malicious activity, especially in a narrow network region. Using SDN in VANETs without any security modifications presents several security challenges since VANET network topology is very dynamic attributed to the mobile nodes [35]. Security systems with AI integration promise to be self-adaptive, affordable, and sustainable even in a changing and unpredictable environment. The proposed AI/ML method used in this work's implementation aims to detect DDoS traffic utilizing an SDN-based VANET as the target system. To effectively assess SDN security, IDS establishes network centrality measurements to identify vulnerable nodes in the network [36].

In this work, we provided an ML approach to security threats in SDN-VANET with a focus on distributed attacks. The proposed RBF-SVM kernel-based framework representing the security framework of SDN may be stored in the SDN controller. The degree of vulnerability of a DDoS attack depends on the configuration and implementation of the SDN controller. The IDS can obtain statistical information about each traffic stream to make decisions to resist a DDOS attack. The detection is carried out by using the traffic sniffing tool and statistics

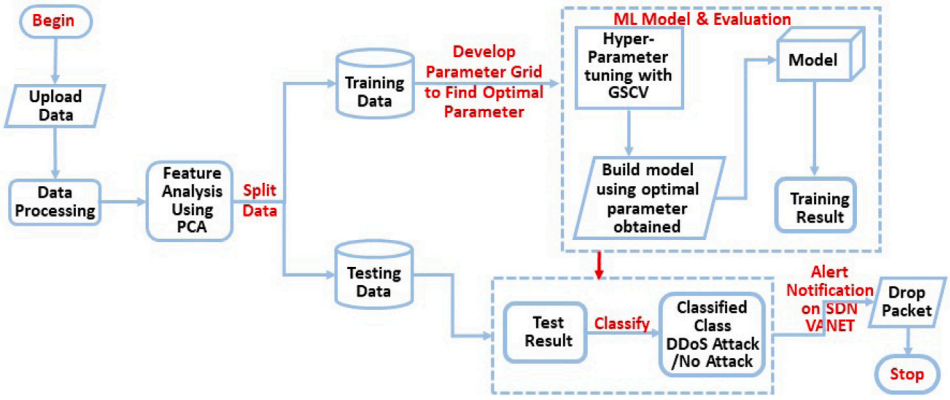


Fig. 2. Flow chart of the proposed RBF-SVM DDoS identification model.

recording feature of the IDS and the Open Flow Switch (OPS) of the SDN controller to analyze the VANET traffic flow features and aggregate those results for the DDOS detection tasks. With the help of the proposed RBF-SVM algorithm, the network packets generated by each device in the data forwarding plane pass through the OPS, which acts as a gateway to the NCP. Once the classification of traffic is done and it detects any attack, an alarm is raised by the OPS. Recall that the SDN controller manages all the OPS in a network.

3.3. Proposed machine learning intrusion detection system for SDN-VANET

ML is a well-known and validated method for predicting, classifying, and detecting attacks among the options for intrusion detection in VANETs. ML-based IDSs employ statistical techniques to characterize typical network behaviors. Any variation in this pattern that exceeds a certain threshold will result in the security system suspecting that an intrusion is in progress and notifying other networked devices. The advantage of this detection method is that it enables the detection of previously unidentified attacks without updating an attack database [37]. An ML-based supervised SVM classifier for DDoS detection and classification is proposed in this study based on the maximum margin classifier. Unquestionably one of the best and most accurate algorithms for classification issues, SVM has the advantage of reaching a global minimum optimality rather than a local minimum, as demonstrated by [38].

As a classification model, an SVM attempts to divide two data subsets into a hyper-plane (decision plane). Each nearest point to the hyper-plane is separated by a maximum margin drawn between these two data subsets. The support vectors are the data points that are closest to the hyperplane of each class [39]. Even in high-dimensional areas with unpredictable data distribution, an IDS can function effectively with SVM. Furthermore, challenging categorization issues can be resolved using the accompanying kernel function. Although previous research has made significant advancements in this field, there is still a need for a further in-depth examination of the impact of kernel type, parameter optimization, and model behavior on dataset features, preprocessing, and characteristics. This work proposed a model that is based on a decision to choose the best hyperplane with near-zero misclassification (a hard margin SVM). Fig. 2 (overall system model) depicts the procedures used to attain the required experimental result and implementation.

3.4. Proposed RBF-SVM model and simulation parameters

The selection of a kernel function and its parameters is crucial because a well-built kernel produces a model that fits well with the underlying data of the structure. Transferring the dataset to a higher-dimensional space is done using the kernel function in SVM [32]. A higher-dimensional space can be created using the SVM kernel functions: linear, polynomial, sigmoid, and RBF. As a result, changing a kernel option enables one to describe how the kernel function affects the model's performance. One of the most thoroughly and permanently smoothed kernels is the RBF Kernel.

RBF kernels can carry out intricate and non-linear mappings more easily than other kernels can, and it is simple to compute parameters, making them ideal for a quick and dependable learning process. When two points are compared, RBF kernels estimate how similar and near they are [40]. Eq. (1) mathematically expresses the RBF kernel ($K_{RBF}(P_1, P_2)$), where σ is the variance and hyper-parameter, d_{12} represents the Euclidean distance $\|P_1 - P_2\|^2$ between two points P_1 and P_2. The symbol exp represents the exponential function.

$$K_{RBF}(P_1, P_2) = exp(-\frac{d_{12}}{2\sigma^2}); \quad (1)$$

3.4.1. Kernel function selection and parameter of proposed model

The RBF kernel function can be specified by adding and adjusting kernel parameters. Proper selection of these parameter values is paramount for the performance of the SVM [38]. While training an RBF-SVM, two parameters must be considered carefully, as they are allied with this kernel. The two parameters associated with the RBF Kernel are 'gamma' (γ) and the penalty parameter ('C'). The parameter 'C' specifies the influence of the kernel on the model and tells the SVM optimization how much misclassification to avoid in each training example. It is common to all SVM kernels and trades off misclassification of training instances for the simplicity of the decision surface [10]. A low 'C' regularizes the decision surface, whereas a high 'C' tries to correctly identify all training samples. The model was optimized using the parameter optimization principle to enhance the accuracy of the default RBF-SVM. The 'γ' parameter indicates how powerful a single training example is.

3.4.2. The grid-search optimization algorithm

Previous work on optimization has provided a basis for optimizing parameters. A wrong choice of parameters may result in a poor fit to the data and, in turn, poor model performance [9]. Hyperparameters are not learned directly within estimators. However, with kernel optimization, a model can yield specific and optimal parameters for the classifier, achieving the lowest error rate while keeping the complexity of the model under control [41]. For an optimal selection of parameters, a cross-validated grid search was implemented on a parameter grid of 'C' and 'γ' to achieve a more optimal selection of parameter values (see Algorithm 1). GSCV is a comprehensive method that searches all possible preset combinations to discover the best point in the domain. Therefore, this optimization method produces a more accurate combination of parameters rather than relying on the randomness implemented in equivalent optimization approaches.

Algorithm 1 K-Fold Cross-Validation with GridSearch.

Ensure: Optimal Hyper-tune parameter C and γ
Require: $RBF - SVM - Accuracy \geq 99\%$
Require: C = 0.01, 0.1, 1, 10, 100, 1000
Require: γ = Scale, 1, 0.1, 0.01, 0.001, 0.0001
▷ C and γ account for the grid point in the grid search
Require: $H_{sets} = C$ and γ combination
Require: D, dataset of DDoS attack features and output binary classification
Require: K_o, K_i, ▷ where K_o is the number of outer folds, and K_i inner folds
for $r, numberof folds = 1 \leftarrow K_o$ splits **do** split D into D_r^{train}, D_r^{test} for the r'th split
end for
for $g, numberof splits = 1 \leftarrow K_i$ splits **do** split D into D_g^{train}, D_g^{test} for the g'th split
end for
for each H in H_{set} **do** train RBF-SVM on D_g^{train}
end for ▷ compute test error E^{test} for RBF-SVM with D_g^{test}
▷ Select optimal hyperparameter set H_o from H_{set}, where E_g^{test} is best
while Train RBF-SVM with D_r^{train}, using H_o **do** Compute test error E_r^{test} for RBF-SVM with D_r^{test}
if $Accuracy \geq 99\%$ **then**
$H_{set} \leftarrow C \times Y$ is optimal
end if
end while

Algorithm 1 shows how to build the GSCV for recognizing DDoS attacks using the proposed method. As a performance metric, a five-fold cross-validation (CV) approach was used, partitioning the D training set into mutually exclusive and exhaustive equal-sized subsets. Fits of K split, five folds for each of the H_{set} 36 possibilities (range of specified 'C' and γ parameter values), for a total of 180 fits. However, the broader the range of parameters, the better the probability of the search mechanism identifying the perfect combination of parameters [9]. Finer adjustment is always possible with a larger range of parameters, but at a considerably higher computational cost. However, computational complexity was taken into account to avoid endless fitting time. Hence, the ranges of 'C' and 'γ' are limited as follows: 'C': [0.01, 0.1, 1, 10, 100, 1000], while 'γ': ["scale", 1, 0.1, 0.01, 0.001, 0.0001].

3.5. Data description

The dataset used for testing and evaluation is the IEEE DataPort [11] for SDN DDoS attacks, which was created to assess DDoS attacks on general SDN-based and Internet of Things (IoT) networks. The SDN DDoS dataset was created using the Mininet emulator and provides an observation of DDoS traffic for identifying suspicious and distributed intrusions in web-connected networks with SDN integration. The development of virtual network topologies is made possible by Mininet, a useful tool. The network simulation runs for 250 min, capturing benign TCP, UDP, and ICMP traffic as well as malicious traffic, which includes TCP, Syn attacks, UDP Flood attacks, and ICMP attacks.

The simulation is carried out for a defined interval using extracted and calculated features. The dataset, consisting of 104,335 samples of 40,784 malignant situations and 63,561 benign scenarios, is translated into CSV (comma separated values) format for the purpose of assessing ML IDSs. The predictors captured from the generated network traffic are outlined in Table 2, overall, there are 22 predictors. The predictors were all taken into account for modeling since they accurately describe the characteristics of SDN-based IoT network traffic. . These are components of the generated network traffic that makes up the independent variables for modeling. The dependent variable is the DDoS attack or no attack label. A detailed description of the extracted predictors and calculated predictors can be found in [42].

Table 2
ML model predictors.

Time_Derivative	Switch_ID
Source_IP	Destination_IP
Packet_Count	Byte_Count
Duration_in_Secs	Duration_in_Nanosecs
Total_Duration	Flows
Packets_ins	PacketperFlow
Byte_per_Flow	Packet_Rate
Pair_Flow	Network_Protocol
Port_Number	Packet_Transmitted (bytes)
Packet_Received (bytes)	Packet_Transmitted (kbps)
Packet_Received (kbps)	Total_Packets (kbps)

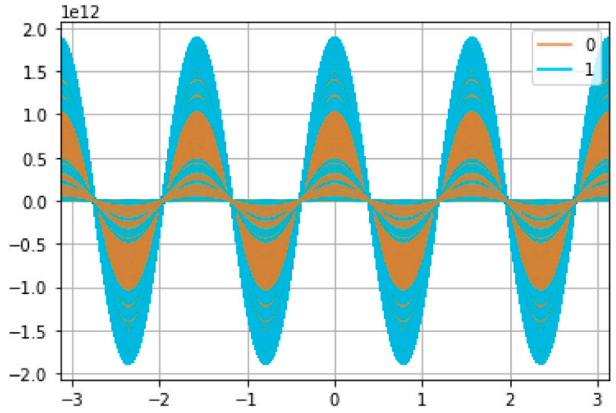


Fig. 3. Andrews plot showing non-linearity of the DDoS dataset.

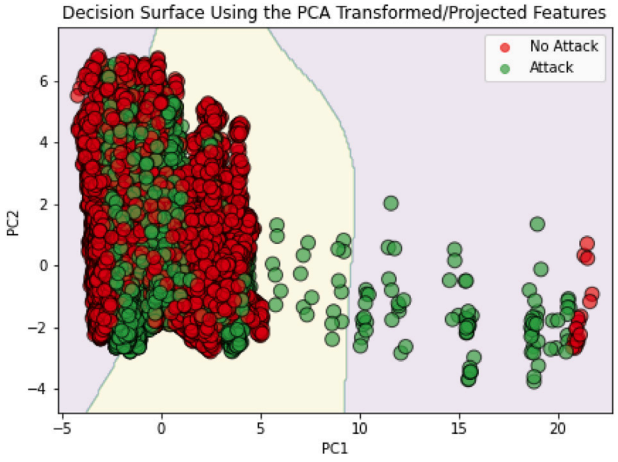


Fig. 4. Decision surface using PCA features.

The Andrews plot Fig. in 3 and PCA plot in Fig. 4 was also employed to visualize clusters and structures in the high-dimensional data showing non-linearity of the data and overlapping class [43]. This confirms that the proposed kernel selection (RBF-SVM) is well suited for modeling.

3.5.1. Principal component metrics and analysis

In a multidimensional investigation, a scree plot represents the eigenvalues of the predictors. When performing a Principal Component Analysis (PCA), the scree plot can visualize how many principal components to preserve [31]. This method aims to discover statistically significant predictors. The explained variances for each predictor are plotted on the scree plot in Fig. 5, sorted by the amount of variation

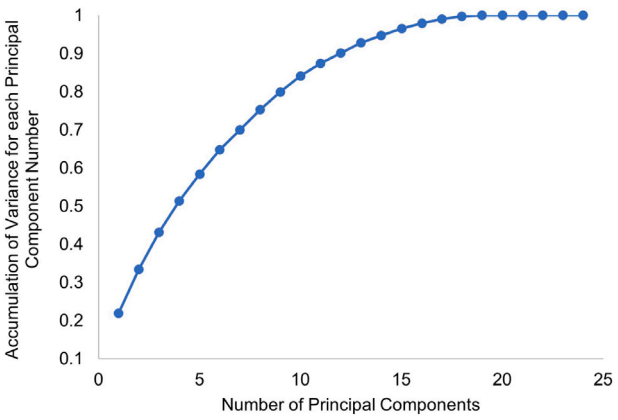


Fig. 5. Scree Plot showing cumulative explained variance of principal components.

each one covers. The feature variances of the dataset are arranged in ascending order, with the eigenvalues going from lowest to highest. This indicates that the final PC represents the eigenvalue that is responsible for the majority of the variance in the information. The next PC had the second most, and so on. In addition, the first 10 components contain approximately over 80% of the variance, while 15 components or more are needed to describe close to 100% of the variance. In conclusion, the majority of predictors possess an explained variance of more than 0.5, confirming their significance in the dataset used for modeling.

3.6. Background information on the state-of-the-art supervised learning algorithms implemented

1. Logistic Regression (LR): LR and SVM generally perform comparably in practice. SVM tries to find the "best" margin that separates classes, reducing the risk of error. Consequently, LR creates decision boundaries with different weights near the optimal point. In addition, both algorithms handle optimization problems indistinguishably [31].
2. K-Nearest Neighbor (KNN): In similarity, kNN and RBF kernel-based SVM are non-parametric methods for estimating the probability density of a given dataset. Both classifiers calculate the distance between instances. A KNN algorithm assigns a new pattern, X, to that category to which the plurality of its k closest neighbors belong. The KNN method can be thought of as estimating the values of the probabilities of the classes given X [37].
3. Gaussian Naive Bayes (GNB): GNB is a Naive Bayes variant that uses the Gaussian normal distribution and works with continuous data. The conditional probability density of response patterns given a stimulus class is modeled using this multivariate classifier. A GNB utilizes a less flexible distributional model, assuming zero off-diagonal co-variance similar to an RBF-SVM [44].
4. Multi-Layer Perceptron (MLP): A fully connected feed-forward ANN having at least three layers—input, output, and at least one hidden layer—is known as MLP. To train MLPs and DNNs, backpropagation can be used [45].

4. Metrics for assessing different techniques in DDoS attacks detection

The performance of the proposed IDS for SDN-based VANET is evaluated through its ability to correctly classify vehicular communication and events as DDoS attacks and benign behaviors. Optimal parameter combination results are obtained and used for training the RBF-SVM kernel method. The chosen evaluation metric is the overall performance and prediction of the proposed model based on the test set, which was 33% of the total data, to see if there is a need for retraining to provide more accurate results. The baseline for performance metrics was an accuracy level higher than 99%. To assess and evaluate the effectiveness of the proposed IDS in comparison with existing state-of-the-art and other ML models, eight (8) performance metrics were considered and outlined below.

1. Accuracy: Accuracy is calculated as the ratio of correctly classified DDoS and normal flow over total flows [31]. It also provides an estimate of how well the IDS can generalize out-of-sample data (test data). Eq. (2) computes the percentage of correct classification.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}; \tag{2}$$

   - True Positive (TP): The result of the model correctly predicting the positive class is TP.
   - True Negative (TN): The result of the model correctly predicting the negative category is TN.
   - False Positive (FP): FP is the result of the misclassified positive class.
   - False Negative (FN): FN represents the prediction of the negative class that does not meet the standards [31].

2. Precision: The proportion of accurately categorized DDoS traffic to all attack records is computed using Eq. (3).

$$Precision = \frac{TP}{TP + FP}; \tag{3}$$

3. Recall: To compute the proportion of accurately detected attacks in the test dataset as a fraction of all attacks [25], Eq. (4) is used.

$$Recall = \frac{TP}{TP + FN}; \tag{4}$$

4. F1 Score: The F1 Score is another widely used metric when employing classification models in ML to evaluate the model's quality.

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall}; \tag{5}$$

5. Matthews Correlation Coefficient (MCC): For binary classification, accuracy may be sensitive to class imbalance while the precision, recall, and F1-score results may seem asymmetric. The MCC is used to compute the correlation between true and predicted values. The higher the correlation between them, the better the prediction. When the classifier is perfect (FP = FN = 0) the value of MCC is 1, indicating perfect positive correlation [46].

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{6}$$

6. R-Squared: R-Squared (R^2) (also known as the coefficient of determination) is a statistical metric that quantifies how much of the variance in the dependent variable can be accounted for by the independent variable. R-squared, thus, displays how well the data fits the model (the goodness of fit). Generally, a higher R^2 indicates more variability is explained by the model [47].

$$R^2 = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y^2)]}} \tag{7}$$

   - n = Number of observations
   - $\sum x$ = Total value of the independent variable
   - $\sum y$ = Total value of the dependent variable
   - $\sum xy$ = Sum of the product of independent and dependent variable

Table 3
Bench-marking the performances of the proposed model against state-of-the-art ML-based IDS implemented in this work.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	MCC (%)	R^2	MAE	CPU time (s)
LR	0.7834	0.7814	0.7834	0.7817	0.5391	0.8842	0.217	141 ms
MLP	0.9571	0.9340	0.9575	0.9456	0.9105	0.8198	0.043	72
KNN	0.9769	0.9741	0.9663	0.9702	0.9515	0.9031	0.023	57
GNB	0.6601	0.5579	0.6070	0.5814	0.2969	−0.4301	0.340	125 ms
Linear SVM	0.8058	0.7684	0.7169	0.7418	0.5875	0.1832	0.194	703
Polynomial-SVM	0.9731	0.9561	0.9755	0.9657	0.9438	0.8868	0.027	78
Sigmoid-SVM	0.5978	0.4832	0.4873	0.4852	0.1554	−0.6919	0.402	280
Default RBF-SVM	0.9769	0.9620	0.9793	0.9706	0.9518	0.9029	0.023	106
Proposed RBF-SVM C: 1000; γ: scale	**0.9940**	**0.9926**	**0.9920**	**0.9923**	**0.9875**	**0.9750**	**0.006**	**74**

- $\sum x^2$ = Sum of the squares of the independent variable value
- $\sum y^2$ = Sum of the squares of the dependent variable value

7. Mean Absolute Error (MAE): The average discrepancy between the observations (actual values) and model output (predictions) calculated using Eq. (8) is known as the MAE. This optimization-based metric is also useful in statistical modeling.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y|; \tag{8}$$

- n = number of data points
- y_i = observed values
- y = predicted values.

8. Resource consumption in terms of the model's execution time (CPU's time was considered): This is the measure of the amount of resources the CPU has spent processing each model. The CPU time is the exact amount of time in seconds, and this forms the basis for measuring how long the model takes to execute.

5. Results and discussion

In this section, we will compare and contrast the proposed RBF-SVM-IDS with the performance of state-of-the-art ML-based IDS implemented in this work, existing ML techniques, SVM kernel variations, RBF parameter selection, dataset property, and scalability. The effectiveness of the proposed model is evaluated using the test set (33%) of the realistic dataset that includes current DDoS attacks in SDN-based situations. The total experimental finding demonstrates that, in comparison to other schemes and arbitrarily chosen parameter schemes, the GSCV-RBF-SVM method may minimize computing complexity while obtaining improved detection accuracy and dependability. Optimal parameter values were captured at C = 1000 and γ = scale respectively. At these optimal parameter points, the model achieved an appreciable and enhanced detection accuracy of **99.40%** in comparison with other state-of-the-art ML approaches, properties, and RBF-SVM structures presented in this work.

5.1. Bench-marking the performances of the proposed model against state-of-the-art ML-based IDS implemented in this work

Four state-of-the-art techniques modeled in this work are LR, MLP, kNN, GNB, including variants of the SVM kernel. The experimental performance and significance of optimizing model parameters are captured in Table 3. The RBF kernel possesses stronger interpolation ability and is good at reflecting local properties of the samples. Its effectiveness on non-linear data is also without doubt. In addition, the complexity associated with this kernel remains the same even with an increase in the input size of the data. The RBF-SVM model was compared with variants of SVM kernels and other ML models for evaluation as shown in Fig. 6. As shown, the proposed RBF-SVM had the best result of 99.4%

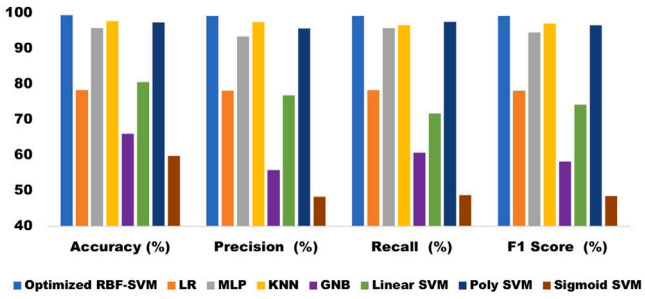


Fig. 6. Proposed model against state-of-the-art ML-based IDS implemented in this work.

accuracy, a recall of 99.20% an F1 score of 99.23%, and a precision of 99.26% executed at a CPU time of 74 s, which is relatively lower compared to other commonly used kernel methods and ML methods.

The interpolation ability of the polynomial kernel is relatively weaker and poorer at reflecting local properties compared to the RBF kernel. However, it can also well extract the global characteristics of the training samples [38]. The 'poly' kernel had the fourth-best classification accuracy of 97.31%, at a precision 95.61%, while the sigmoid kernel had the least experimental performance even after data conversion into less dimensional space. Linear kernel SVM models work best with small-dimensional data elements arranged in a sequential manner [38]. High-dimensional data makes linear computation more complex, consuming more computational resources. The linear kernel is not so effective on datasets with overlapping classes, as shown in the Andrews plot in Fig. 3, hence it is impossible to generate results within a reasonable time. Hence, the highest computational resource of 703 s was recorded by the linear kernel with an accuracy of 80.58%.

5.2. Comparison under various RBF-SVM kernel parameter conditions

To distinguish between default, random, and optimal parameter selection, an additional evaluation methodology was carried out by testing several RBF-SVM variants. Table 4 provides a tabular representation of results for the SDN DDoS dataset, comparing the proposed model with parameter performances for a variety of random parameter values. In all cases, hyperparameters were selected randomly. The combination of RBF-SVM and GSCV as optimization techniques achieved better accuracy performance. The accuracy achieved by the proposed IDS confirms the necessity for implementing the hyper-parameter optimization and selection technique rather than random parameter selections.

To extensively validate the proposed model's results, other performance metrics for evaluating reliability and error rate introduced in the previous section were also captured in this table. Following closely the proposed model in terms of classification performance was the RBF-kernel variant with C = 10 00 000 and γ as 0.0046, having an accuracy of 99.36%, a closely accurate recall, F1 score, and precision, although a longer computing time of 3054 s was recorded. The same trend of increased computing resources is observed in other parameter variants. Experimental results of C = 100 and γ = 0.0001 recorded the poorest performance with a very poor accuracy of 86.47%.

Table 4
Bench-marking other various parameter conditions.

RBF-SVM parameter variants	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	MCC (%)	R^2	MAE	CPU time (s)
Default RBF-SVM	0.9769	0.9620	0.9793	0.9706	0.9518	0.9029	0.023	106
RBF-SVM C: 1 000 000; γ: 0.0046	0.9936	0.9924	0.9913	0.9918	0.9867	0.9734	0.006	3054
RBF-SVM C: 1000; γ: 0.001	0.9686	0.9445	0.9767	0.9603	0.9348	0.8681	0.031	287
RBF-SVM C = 100, γ = 0.0001	0.8647	0.8438	0.8002	0.8214	0.7134	0.4308	0.135	340
RBF-SVM C = 100, γ = 0.01	0.9823	0.9720	0.9828	0.9773	0.9629	0.9255	0.018	95
RBF-SVM C = 100, γ = 0.001	0.9573	0.9238	0.9704	0.9465	0.9119	0.8207	0.043	141
RBF-SVM C = 100, γ = scale	0.9922	0.9880	0.9921	0.9900	0.9838	0.9675	0.008	81
RBF-SVM C = 100, γ = 1	0.9801	0.9812	0.9674	0.9743	0.9582	0.9165	0.020	1641
Proposed RBF-SVM C: 1000; γ: scale	**0.9940**	**0.9926**	**0.9920**	**0.9923**	**0.9875**	**0.9750**	**0.006**	**74**

5.3. Matthews correlation coefficient evaluation

Among the most widely used measures for binary classification problems are accuracy, recall, and F1 score, all calculated on confusion matrices. However, these statistical techniques, particularly on unbalanced datasets, can dangerously reveal overly optimistic or exaggerated outcomes. Instead, the Matthews correlation coefficient (MCC) is a more dependable statistical measure that only yields a high score if the prediction performed well in each of the four categories of the confusion matrix (TP, FP, FN, and FP), proportionally to the size of the dataset's positive and negative elements [46]. The high MCC (**0.9875**) of the proposed model in contrast with others indicates that the majority of positive data instances and the majority of negative data instances could be reliably predicted by the proposed binary predictor.

5.4. Coefficient of Determination (R^2) Evaluation

The lowest possible value of R^2 is 0 and the highest possible value is 1. Put simply, the better a model is at making predictions, the closer its R^2 will be to 1. A model's performance can be evaluated using the R^2 by graphing it. Fig. 7 display two sets of data (predicted and actual) capturing the majority of points around the line of best fit. The blue line represents the actual data while the black dashed lines represent the predicted data, which also denotes the line of best fit, or predictions of the model. It can be observed that only a few deviations are captured in the plot. The few blue dots represent the deviations (the residuals) between the actual data and their predicted values. This demonstrates how well the observations match the model's predictions, as the model produced a high R^2 of 0.9750.

5.5. MAE evaluation

The MAE score, which measures the average of the absolute error values, provides an intuitive error value match against the predicted target value units. An MAE measures how far 'off' a measurement is from a true value or an indication of the uncertainty in a measurement. Here we compare the results in terms of MAE value among existing methods and the proposed method by analyzing the SDN-DDoS dataset. As we obtained a lower MAE value of **0.006** for our proposed approach,

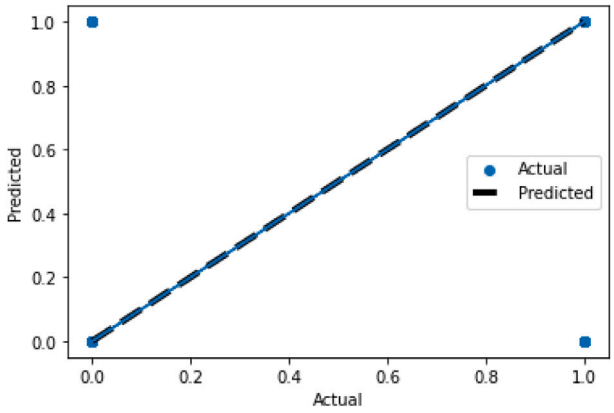


Fig. 7. Coefficient of determination of proposed RBF-SVM.

we can say that our approach is a reliable DDoS detection technique. In the MAE column 8th, column of Tables 3 and 4, MAE values of all models are shown for the SDN-DDoS dataset. The lower the MAE value, the higher the accuracy. By using the optimized RBF-SVM technique, we achieve lower MAE values and high accuracy when compared to all approaches.

5.6. Resource consumption in terms of the proposed model's execution time

In real-time applications, execution time plays an important role, and so does classification performance. Therefore, computational analyses relying on the models' execution time were performed. The execution durations are presented in the last columns of Tables 3 and 4. All experimental studies were carried out in a Jupyter Python environment running Windows 10 operating system on a computer with an Intel(R) Core (TM) i5-8500 CPU @ 3.00 GHZ with 8 GB RAM, in all experiments. According to these results, the proposed classifier at (**74 s**) yielded the lowest total execution time among all models. In conformity with these results, it is concluded that a faster detection system is obtained using the optimal parameters. In this study, the hyper-parameter

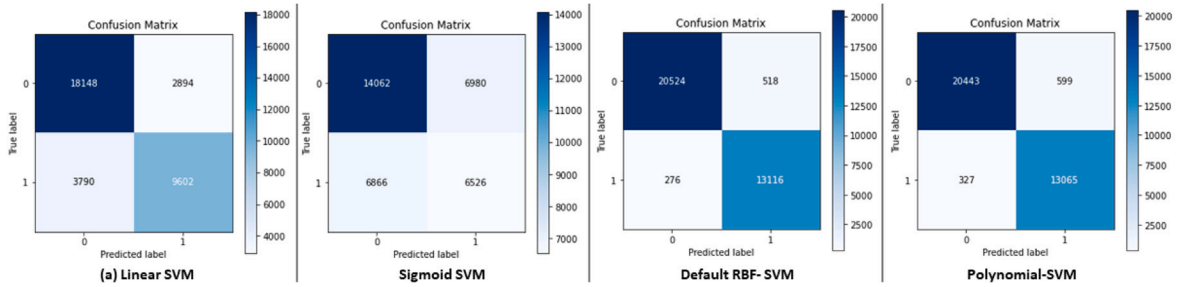


Fig. 8. Confusion matrix of SVM kernel methods (8a) Linear SVM, (8b) Sigmoid SVM, (8c) Default RBF-SVM and (8d) Polynomial SVM respectively.

optimization for the proposed ML classifier was carried out using the GSCV optimization method.

5.7. Benchmarking the performances of the proposed model against state-of-the-art ML-based IDS

The accuracy results obtained using our proposed approach when compared with related studies carried out in the literature are presented in the 8th column of Table 1. Comparing the performance of the proposed RBF-SVM-based algorithm to comparable supervised and traditional techniques from related studies validates the effectiveness of the proposed algorithm. The suggested IDS had the highest accuracy, whereas the SVM model employed by the authors in [14] had the second-best performance of 98.59%. The third-best is recorded as 98.07% by authors in [20], with an accuracy of 97.88%, KNN comes in third place. The least accurate model (86.63%) [25]. The selected SVM approach outperformed all SVM methodologies used in related works [15–18,20–25] with an accuracy of 99.40%. For a mission-critical system, accuracy levels of over 99% are usually required.

5.8. Mis-classification results using confusion matrix

A confusion matrix is a two-dimensional array that compares predicted category labels to the actual label in binary classifications. To establish how many observations were successfully or wrongly categorized, the performance of the proposed model was represented using a confusion matrix. The confusion matrix of the SVM kernel variants (linear SVM, sigmoid SVM, default RBF-SVM, and polynomial SVM, respectively) is presented in Fig. 8. The highest misclassification rate was observed using the sigmoid kernel. Though fewer incorrect classifications occurred using the default RBF, the proposed model produced superior classification accuracy and much lower misclassifications. The number of accurate and inaccurate predictions made by the default RBF-SVM kernel displayed in Fig. 9 in contrast with that of the proposed RBF-SVM version Fig. 10 shows fewer misclassifications.

5.9. Evaluating the proposed model's scalability on larger DDoS dataset

For further evaluation, the CICDDoS2019 dataset was also considered [48]. The CICDDoS2019 dataset captures reflection-based and exploitation-based attack types and is a representative for DDoS evaluation in any network. The dataset contains raw data, including network traffic (Pcaps) and event logs. CICFlowMeter-V3 was used to extract more than 80 traffic features. In both datasets, each DDoS attack instance is labeled as '1' while normal data communication has its behavior set as '0'. The data is preprocessed to handle missing and categorical attributes before training and testing. Disregarding this technique could thwart the ability of the IDS to correctly classify all instances of benign and malignant data [37]. Further dataset exploratory techniques were also performed to increase its interpretability and minimize information loss.

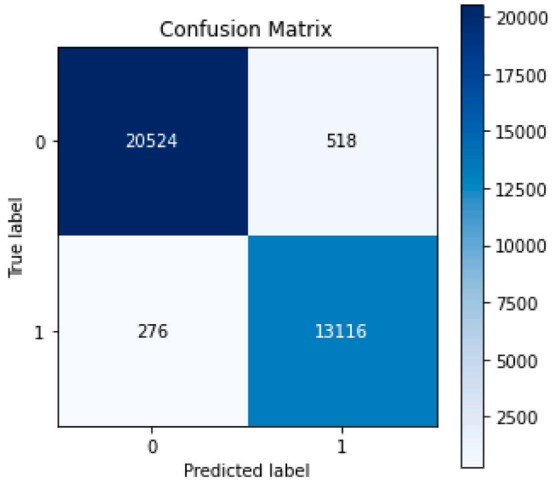


Fig. 9. Confusion matrix of default RBF-SVM showing TP, TN, FN and FP.

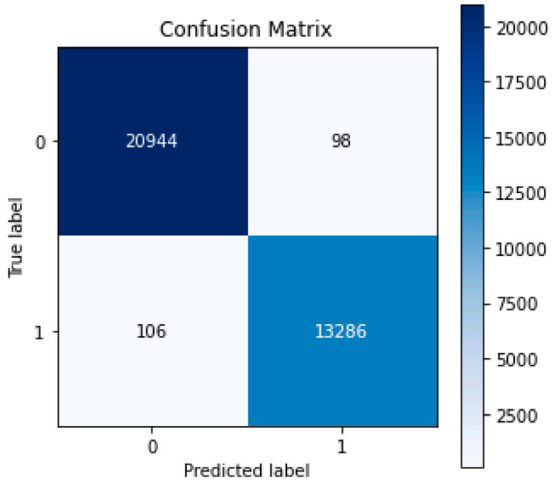


Fig. 10. Confusion matrix of proposed RBF-SVM showing TP, TN, FN and FP.

An additional thorough analysis of the proposed model's impact on a more widespread type of DDoS attacks was explored using the CICDDoS2019 dataset. The CICDDoS2019 dataset is apt for assessment, security testing, and malware prevention in huge networks [48]. The proposed RBF-based kernel utilizes computational memory very effectively. The complexity of the suggested kernel is constant regardless of the amount of data input. The experimental outcome of the proposed model on the CICDDoS2019 dataset is displayed in Table 5, reaching an accuracy of 99.96%. With a 73 s computation time, the model produced results with a precision of 98.46%, an f1 score of 99.22 and a recall of 100%.

Table 5
Evaluating the scalability of the proposed model on the CICDDoS2019 dataset.

Accuracy (%)	99.96
Precision (%)	98.46
Recall (%)	100
F1 Score (%)	99.22
MCC (%)	99.21
R^2	0.9840
MAE	0.006
CPU time	73 s

6. Conclusion

An IDS for SDN-integrated VANETs is proposed in this study. Based on the SDN DDoS attack dataset, the proposed model was evaluated. To minimize absolute error and obtain better accuracy while minimizing complexity, our model incorporated parameter optimization. The GSCV technique was used for optimization to curb overfitting and achieve good performance. The combination of the RBF-SVM kernel and the GSCV method achieved good performance. Several evaluation metrics were used to assess the performance of the various ML methods compared in this work. The experimental results show that it is possible to detect attacks with very high accuracy. The proposed model had the best performance of 99.40% accuracy and 99.26% precision in binary classification problems compared to all other ML methods. Additionally, the framework can benefit from different types of attack detection, using more complex data at higher levels.

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.

Data availability

Data will be made available on request.

Acknowledgment

This research was supported by Kumoh National Institute of Technology (202001690001).

References

[1] D. Javeed, T. Gao, M.T. Khan, SDN-enabled hybrid DL-driven framework for the detection of emerging cyber threats in IoT, Electronics 10 (8) (2021) http://dx.doi.org/10.3390/electronics10080918.

[2] F. Gonçalves, J. Macedo, A. Santos, An intelligent hierarchical security framework for VANETs, Information 12 (11) (2021) http://dx.doi.org/10.3390/info12110455.

[3] J.C. Nobre, A.M. de Souza, D. Rosário, C. Both, L.A. Villas, E. Cerqueira, T. Braun, M. Gerla, Vehicular software-defined networking and fog computing: Integration and design principles, Ad Hoc Netw. 82 (2019) 172–181, http://dx.doi.org/10.1016/j.adhoc.2018.07.016.

[4] G.C. Amaizu, C.I. Nwakanma, S. Bhardwaj, J.-M. Lee, D.-S. Kim, Composite and efficient ddos attack detection framework for B5G networks, Comput. Netw. 188 (2021) 107871, http://dx.doi.org/10.1016/j.comnet.2021.107871.

[5] A. Zainudin, L.A.C. Ahakonye, R. Akter, D.-S. Kim, J.-M. Lee, An efficient hybrid-DNN for ddos detection and classification in software-defined iIoT networks, IEEE Internet Things J. (2022) 1, http://dx.doi.org/10.1109/JIOT.2022.3196942.

[6] M. Dibaei, X. Zheng, K. Jiang, R. Abbas, S. Liu, Y. Zhang, Y. Xiang, S. Yu, Attacks and defences on intelligent connected vehicles: A survey, Digit. Commun. Netw. 6 (4) (2020) 399–421, http://dx.doi.org/10.1016/j.dcan.2020.04.007.

[7] M. AbdelBasset, N. Moustafa, H. Hawash, W. Ding, Internet of things security requirements, threats, attacks, and countermeasures, in: Deep Learning Techniques for IoT Security and Privacy, Springer International Publishing, Cham, 2022, pp. 67–112, http://dx.doi.org/10.1007/978-3-030-89025-4_3.

[8] G.O. Anyanwu, C.I. Nwakanma, J.-M. Lee, D.-S. Kim, Real-time position falsification attack detection system for internet of vehicles, in: 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2021, pp. 1–4, http://dx.doi.org/10.1109/ETFA45728.2021.9613271.

[9] I. Syarif, A. Prügel-Bennett, G.B. Wills, SVM parameter optimization using grid search and genetic algorithm to improve classification performance, TELKOMNIKA Telecommun. Comput. Electron. Control 14 (2016) 1502–1509, http://dx.doi.org/10.12928/telkomnika.v14i4.3956.

[10] A. Rahimi, B. Recht, Random features for large-scale kernel machines, in: J. Platt, D. Koller, Y. Singer, S. Roweis (Eds.), Advances in Neural Information Processing Systems, 20, Curran Associates, Inc., 2007, pp. 1177–1184.

[11] S. Sambangi, L. Gondi, S. Aljawarneh, S.R. Annaluri, SDN DDOS attack image dataset, 2021, http://dx.doi.org/10.21227/k06q-3t33.

[12] J. Asharf, N. Moustafa, H. Khurshid, E. Debie, W. Haider, A. Wahab, A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions, Electronics 9 (7) (2020) http://dx.doi.org/10.3390/electronics9071177.

[13] S. Campos-Cordobés, J. del Ser, I. Laña, I.I. Olabarrieta, J. Sánchez-Cubillo, J.J. Sánchez-Medina, A.I. Torre-Bastida, Chapter 5 - big data in road transport and mobility research, in: F. Jiménez (Ed.), Intelligent Vehicles: Enabling Technologies and Future Developments, Butterworth-Heinemann, 2018, pp. 175–205, http://dx.doi.org/10.1016/B978-0-12-812800-8.00005-9.

[14] O. Tonkal, H. Polat, E. Baåaran, Z. Cömert, R. Kocaoğlu, Machine learning approach equipped with neighbourhood component analysis for DDoS attack detection in software-defined networking, Electronics 10 (11) (2021) http://dx.doi.org/10.3390/electronics10111227.

[15] P.K. Singh, S. Kumar Jha, S.K. Nandi, S. Nandi, ML-based approach to detect ddos attack in V2I communication under sdn architecture, in: TENCON 2018 - 2018 IEEE Region 10 Conference, 2018, pp. 0144–0149, http://dx.doi.org/10.1109/TENCON.2018.8650452.

[16] H. Polat, M. Turkoglu, O. Polat, Deep network approach with stacked sparse autoencoders in detection of ddos attacks on SDN-based VANET, IET Commun. 14 (2020) 4089–4100, (11).

[17] F.A. Alhaidari, A.M. Alrehan, A simulation work for generating a novel dataset to detect distributed denial of service attacks on vehicular ad hoc network systems, Int. J. Distrib. Sens. Netw. 17 (3) (2021) http://dx.doi.org/10.1177/15501477211000287.

[18] M. Türkoğlu, H. Polat, C. Koçak, O. Polat, Recognition of ddos attacks on SD-VANET based on combination of hyperparameter optimization and feature selection, Expert Syst. Appl. 203 (2022) 117500, http://dx.doi.org/10.1016/j.eswa.2022.117500.

[19] K. Savitha, D.C. Chandrasekar, A hybrid intrusion detection model for VANET using SDN and growing hierarchical self-organizing maps, in: Webology, Vol. 18, 2022, pp. 158–182.

[20] K.S. Sahoo, B.K. Tripathy, K. Naik, S. Ramasubbareddy, B. Balusamy, M. Khari, D. Burgos, An evolutionary SVM model for DDOS attack detection in software defined networks, IEEE Access 8 (2020) 132502–132513, http://dx.doi.org/10.1109/ACCESS.2020.3009733.

[21] Y. Yu, L. Guo, Y. Liu, J. Zheng, Y. Zong, An efficient SDN-based ddos attack detection and rapid response platform in vehicular networks, IEEE Access 6 (2018) 44570–44579, http://dx.doi.org/10.1109/ACCESS.2018.2854567.

[22] K. Adhikary, S. Bhushan, S. Kumar, K. Dutta, Hybrid algorithm to detect DDoS attacks in VANETs, Wirel. Pers. Commun. 114 (2020) 3613–3634, http://dx.doi.org/10.1007/s11277-020-07549-y.

[23] A. Kaushik, B. Shashi, K. Sunil, D. Kamlesh, Decision tree and neural network based hybrid algorithm for detecting and preventing DDoS attacks in VANETS, Int. J. Innov. Technol. Explor. Eng. (IJITEE) 9 (2020) http://dx.doi.org/10.35940/ijitee.E2652.039520.

[24] J.A. Pérez-Díaz, I.A. Valdovinos, K.-K.R. Choo, D. Zhu, A flexible SDN-based architecture for identifying and mitigating low-rate DDoS attacks using machine learning, IEEE Access 8 (2020) 155859–155872, http://dx.doi.org/10.1109/ACCESS.2020.3019330.

[25] Y. Gao, H. Wu, B. Song, Y. Jin, X. Luo, X. Zeng, A distributed network intrusion detection system for distributed denial of service attacks in vehicular ad hoc network, IEEE Access 7 (2019) 154560–154571, http://dx.doi.org/10.1109/ACCESS.2019.2948382.

[26] G. Kaur, D. Kakkar, Hybrid optimization enabled trust-based secure routing with deep learning-based attack detection in VANET, Ad Hoc Netw. 136 (2022) 102961, http://dx.doi.org/10.1016/j.adhoc.2022.102961.

[27] M. Aloqaily, S. Otoum, I.A. Ridhawi, Y. Jararweh, An intrusion detection system for connected vehicles in smart cities, Ad Hoc Netw. 90 (2019) 101842, http://dx.doi.org/10.1016/j.adhoc.2019.02.001, Recent advances on security and privacy in Intelligent Transportation Systems.

[28] M. Adnan, J. Iqbal, A. Waheed, N.U. Amin, M. Zareei, A. Umer, E.M. Mohamed, Towards the design of efficient and secure architecture for software-defined vehicular networks, Sensors 21 (11) (2021) http://dx.doi.org/10.3390/s21113902.

[29] M.S. Todorova, S.T. Todorova, DDoS attack detection in SDN-based VANET architectures, 2016, p. 175, URL https://projekter.aau.dk/projekter/en/student-thesis/ddos-attack-detection-in-sdnbased-vanet-architectures.html.
[30] M. Safwat, A. Elgammal, E.G. AbdAllah, M.A. Azer, Survey and taxonomy of information-centric vehicular networking security attacks, Ad Hoc Netw. 124 (2022) 102696, http://dx.doi.org/10.1016/j.adhoc.2021.102696.
[31] A. Subasi, Chapter 3 - machine learning techniques, in: A. Subasi (Ed.), Practical Machine Learning for Data Analysis using Python, Academic Press, 2020, pp. 91–202, http://dx.doi.org/10.1016/B978-0-12-821379-7.00003-5.
[32] S. Tong, D. Koller, Support vector machine active learning with applications to text classification, J. Mach. Learn. Res. (2001) 45–66.
[33] F. Bensalah, N. Elkamoun, Y. Baddi, SDNStat-sec: A statistical defense mechanism against ddos attacks in SDN-based VANET, in: F. Saeed, T. Al-Hadhrami, F. Mohammed, E. Mohammed (Eds.), Advances on Smart and Soft Computing, Springer Singapore, Singapore, 2021, pp. 527–540.
[34] R.F. Fouladi, O. Ermiå, E. Anarim, A DDoS attack detection and countermeasure scheme based on DWT and auto-encoder neural network for SDN, Comput. Netw. 214 (2022) 109140, http://dx.doi.org/10.1016/j.comnet.2022.109140.
[35] W. Ben Jaballah, M. Conti, C. Lal, Security and design requirements for software-defined VANETs, Comput. Netw. 169 (2020) 107099, http://dx.doi.org/10.1016/j.comnet.2020.107099.
[36] T. Eom, J.B. Hong, S. An, J.S. Park, D.S. Kim, A systematic approach to threat modeling and security analysis for software defined networking, IEEE Access 7 (2019) 137432–137445, http://dx.doi.org/10.1109/ACCESS.2019.2940039.
[37] A. Subasi, Chapter 2 - data preprocessing, in: A. Subasi (Ed.), Practical Machine Learning for Data Analysis using Python, Academic Press, 2020, pp. 27–89, http://dx.doi.org/10.1016/B978-0-12-821379-7.00002-3.
[38] K. El Bouchefry, R.S. de Souza, Chapter 12 - learning in big data: Introduction to machine learning, in: P. Åkoda, F. Adam (Eds.), Knowledge Discovery in Big Data from Astronomy and Earth Observation, Elsevier, 2020, pp. 225–249, http://dx.doi.org/10.1016/B978-0-12-819154-5.00023-0.
[39] S.M. Basha, D.S. Rajput, Chapter 9 - survey on evaluating the performance of machine learning algorithms: Past contributions and future roadmap, in: A.K. Sangaiah (Ed.), Deep Learning and Parallel Computing Environment for Bioengineering Systems, Academic Press, 2019, pp. 153–164, http://dx.doi.org/10.1016/B978-0-12-816718-2.00016-6.
[40] P. Chudzian, Radial basis function kernel optimization for pattern classification, in: R. Burduk, M. Kurzyåski, M. Woåniak, A. Åoånierek (Eds.), Computer Recognition Systems. Advances in Intelligent and Soft Computing, Vol. 95, Springer, Berlin, Heidelberg, 2011, pp. 99–108, http://dx.doi.org/10.1007/978-3-642-20320-6-11.
[41] T. Yu, H. Zhu, Hyper-parameter optimization: A review of algorithms and applications, 2020, arXiv:2003.05689.
[42] N. Ahuja, G. Singal, D. Mukhopadhyay, DDOS attack SDN dataset, 2020, http://dx.doi.org/10.17632/jxpfjc64kr.1.
[43] V. Grinshpun, Application of Andrew's plots to visualization of multidimensional data, Int. J. Environ. Sci. Educ. 11 (17) (2016) 10539–10551, http://dx.doi.org/10.3390/info10030106.
[44] M. Misaki, Y. Kim, P.A. Bandettini, N. Kriegeskorte, Comparison of multivariate classifiers and response normalizations for pattern-information fMRI, NeuroImage 53 (1) (2010) 103–118, http://dx.doi.org/10.1016/j.neuroimage.2010.05.051.
[45] D. Sarkar, R. Bali, T. Sharma, Practical machine learning with python, in: A Problem-Solver's Guide to Building Real-World Intelligent Systems, Apress Berkeley, CA, 2018, p. 530, http://dx.doi.org/10.1007/978-1-4842-3207-1.
[46] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (MCC) over F1-score and accuracy in binary classification evaluation, BMC Genomics 21 (6) (2020) http://dx.doi.org/10.1186/s12864-019-6413-7.
[47] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ. Comput. Sci. 7 (e623.) (2021) http://dx.doi.org/10.7717/peerj-cs.623.
[48] I. Sharafaldin, A.H. Lashkari, S. Hakak, A.A. Ghorbani, Developing realistic distributed denial of service (ddos) attack dataset and taxonomy, in: 2019 International Carnahan Conference on Security Technology (ICCST), 2019, pp. 1–8, http://dx.doi.org/10.1109/CCST.2019.8888419.

Goodness Oluchi Anyanwu graduated with a B. Tech degree in Information Management Technology from the Federal University of Technology, Owerri, Imo State Nigeria in 2018. She is currently a Master's Student in Information Technology Convergence Engineering and a Full Time Researcher at Networked System Laboratory (NSL), Kumoh National Institute of Technology, (KIT), Gumi, South Korea.

Cosmas Ifeanyi Nwakanma (Member, IEEE) received the Ph.D. degree in IT-Convergence Engineering from the Networked System Laboratory, IT-Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea in 2022. He received the Diploma (Distinction) in Electrical/Electronics Engineering from Federal Polytechnic Nekede Imo State Nigeria in 1999, and the Bachelor of Engineering degree in communication engineering, the master's degree in information technology, and the Master of Business Administration (MBA) degree in project management technology from the Federal University of Technology, Owerri, in 2004, 2012, and 2016, respectively. Since 2009, he has been lecturing and doing research at the Federal University of Technology, Owerri. He was an intern with Asea Brown Boveri (ABB), Nigeria in 2003. He is also a postdoctoral researcher at the ICT Convergence Research Center (ICTCRC), Kumoh National Institute of Technology, Gumi, South Korea. He is a member of IEEE, Computer Professionals council of Nigeria (CPN), Nigeria Society of Engineers (NSE) and registered by the Council for the Regulation of Engineering in Nigeria (COREN). His research interests include reliability of artificial intelligence (AI) application to Internet of Things (IoT) for smart factories, homes, vehicles and Metaverse.

Jae-Min Lee (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2005. From 2005 to 2014, he was a Senior Engineer with Samsung Electronics, Suwon, South Korea. From 2015 to 2016, he was a Principal Engineer at Samsung Electronics. Since 2017, he has been an Assistant Professor with the Department of IT-Convergence Engineering, School of Electronic Engineering, Kumoh National Institute of Technology, Gyeongbuk, South Korea. His current research interests include industrial wireless control networks, performance analysis of wireless networks, and TRIZ.

Dong-Seong Kim (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2003. From 1994 to 2003, he worked as a full time Researcher at ERC-ACI, Seoul National University. From March 2003 to February 2005, he worked as a Postdoctoral Researcher at the Wireless Network Laboratory, School of Electrical and Computer Engineering, Cornell University, NY, USA. From 2007 to 2009, he was a Visiting Professor with the Department of Computer Science, University of California, Davis, CA, USA. He is currently the Director of the KIT Convergence Research Institute and ICT Convergence Research Center (ITRC and NRF Advanced Research Center Program) supported by Korean Government, Kumoh National Institute of Technology. His research interests include real-time IoT and smart platform, industrial wireless control networks, and networked embedded systems. He is a Senior Member of ACM.