International Journal of Safety and Security Engineering
Vol. 14, No. 5, October, 2024, pp. 1517-1525
Journal homepage: http://iieta.org/journals/ijsse

Securing Smart Grids: Machine Learning-Driven Ensemble Intrusion Detection for IoT RPL Networks

Omar A. Abdulkareem[1,2*], Raja Kumar Kontham[1], Farhad E. Mahmood[3]

[1] Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam 530003, India
[2] Directorate of Research and Development, Ministry of Higher Education and Scientific Research, Baghdad 10065, Iraq
[3] Department of Electrical Engineering, University of Mosul, Mosul 41002, Iraq

Corresponding Author Email: omarasalam.rs@andhrauniversity.edu.in

https://doi.org/10.18280/ijsse.140519

Received: 21 May 2024
Revised: 11 July 2024
Accepted: 29 July 2024
Available online: 31 October 2024

Keywords:
Internet of Things (IoT), routing protocol (RPL), intrusion detection system (IDS), version number (VN), hello flood (HF), decrease rank (DR), smart grid, 5-fold cross-validation, CrowdStrike

ABSTRACT

As the Internet of Things (IoT) continues to expand in the industrial domain, cyber threats have become a major concern, with routing attacks posing significant risks due to the heterogeneity of IoT devices, limited resources, and extensive connectivity. This research aims to enhance the security of IoT-based RPL networks by developing an advanced Intrusion Detection System (IDS) utilizing ensemble learning techniques. The primary objective is to create a robust cybersecurity solution capable of detecting and mitigating Version Number (VN), Hello Flood (HF), and Decrease Rank (DR) attacks, which can cause substantial disruptions and data loss. The proposed IDS model is validated using the IRAD dataset, attaining exceptional performance with 99.88% accuracy, precision, recall, and F1 scores. The methodology incorporates a 5-fold cross-validation approach to confirm reliability and scalability. Comparative analysis with existing models validates the statistical significance and robustness of the proposed solution, highlighting its effectiveness in enhancing IoT network security against evolving cyber threats. This study underscores the critical need for advanced IDS solutions to safeguard the integrity and functionality of IoT networks. Additionally, recent incidents, such as the CrowdStrike 2024 incident, highlight the ongoing challenge and the importance of robust cybersecurity measures in today's digital landscape.

1. INTRODUCTION

The Internet of Things (IoT) is a rapidly emerging technology that has the potential to revolutionize many aspects of our lives. IoT denotes a network of physical devices equipped with sensors, radio communication, software, and other components. These devices collectively enable the acquisition, processing, and bidirectional transfer of data. These generated data are useful when utilized to enhance our understanding of the global environment and to develop more sustainable solutions. For example, IoT-enabled devices can be used to supervise and control our homes, businesses, and transportation systems. They can also provide real-time data on environmental conditions, such as weather monitoring and water pollution. These data can be utilized to improve our understanding of the environment and to develop more sustainable solutions. Moreover, The International Data Corporation a worldwide market intelligence provider, predicts that an increase in the tally of IoT-connected things, such as (actuators, sensors, cameras, etc.) connected online will be 41.7 billion by 2024, which will generate around 79.4 zetta-bytes [1].

The importance of robust cybersecurity measures is further underscored by recent events such as the 2024 CrowdStrike incident, where adversaries leveraged stolen identity credentials to exploit gaps in cloud environments, emphasizing the need for advanced detection and response strategies. This incident highlighted a dramatic increase in attack velocity and the exploitation of generative AI to reduce the barrier of entry for more complicated operations, reinforcing the need for comprehensive security frameworks such as the one we propose in this study.

1.1 Literature review

The proliferation of these connected things can be largely attributed to the cost-effectiveness of IoT devices and related components, in addition to the growing demand for data-driven decision-making. IoT applications have a huge impact on our daily lives. Regardless of their benefits, they are also risky and put the user's security and privacy in danger [2]. In the domain of low-power wireless communication, the demand.

We need a push to develop an efficient routing protocol called 6LoWPAN [3]. Thus, to overcome these limitations, the Internet Engineering Task Force (IETF) researched to determine the practicability of using standard routing protocols to manage the data flow between nodes within the low-power wide-area network (LLNs) and a central hub in the LLN network. The IETF Working Group concluded that these

conventional protocols were unsatisfactory in meeting the distinct routing needs of LLNs. Hence, they introduced and standardized what is called the IPV6 routing protocol for Low Power Lossy Network protocol (RPL) as an efficient solution to manage routing traffic within LLNs networks [4].

RPL is developed specifically to be used in IoT networks, including smart cities, smart grids, healthcare, and machine to machine networks, addressing their routing needs and resource efficiency. Therefore, it is vital to examine the security aspects of RPL in order to gain a better awareness of attacks and mitigation strategies for RPL [5]. Because of the limited energy and constraints of smart devices (Computing power, Storage), traditional security countermeasures like cryptography cannot be applied to secure IoT networks. Therefore, alternative solutions are needed to protect devices from attacks [6].

Figure 1 illustrates the operation of a smart grid network using the RPL protocol, which consists of IoT devices and smart grid devices, protected by intrusion detection systems against attacks that target the RPL-based smart grid network.

In this study, we aim to protect the RPL network by proposing an intrusion detection system (IDS)-based ML that can defend against three RPL routing-specific attacks, namely VN, DR, and HF. The system is based on an ensemble learning technique using four algorithms, i.e., DT, RF, and ET as a based learner and then their prediction feeds to the meta learner. We evaluated the model using the IRAD dataset [7] and conducted an assessment of classifier performance by utilizing a variety of evaluation metrics.

1.2 Attacks on the RPL of smart grid

Many variations of attacks target RPL networks, some of which stem from the sensor network, while others are specific to RPL and exploit its functionality. RPL networks are vulnerable to both passive and active attacks [8]. Researchers classified attacks into three categories of security attacks. The first group is attacks that deplete network resources (energy, memory, and power). The second category of RPL attacks is attacks that target the network's topology by disturbing the routing information carried in RPL control messages. This disruption may lead to routing loops, packet loss, and other issues. The third type of attack aims to interfere with the transmission of data within a network, for example, by listening in or intercepting packets [9]. The primary objective of this study is to identify three attacks that target the RPL network, namely VN, DR, and HF, which are explained below.

1.2.1 Hello flood attacks (HF)

In the context of network security, the threat of a malicious node assumes significance because it may adopt the guise of a new node, periodically dispatching DIS control messages to neighboring nodes. Consequently, nodes situated within the proximity of the adversarial entity are compelled to either reset their Trickle timers or transmit DIS messages to a designated node, thereby necessitating a subsequent delivery of DIO as a response. This sequence of events may result in the congestion of the RPL network nodes, which is chiefly attributable to the heightened volume of routing control [10]. It has been demonstrated that HFA is the most significant attack that adversely affects the performance of the IoT network [11].

1.2.2 Version number attacks (VN)

The attack in question leverages a vulnerability in the global repair function of the RPL. Notably, by RPL's design, the modification of the (DODAG) version number field of the DIO control message is an exclusive prerogative of the root node, with other nodes intended to maintain the version number as is. The current version of the RPL standard lacks a dedicated mechanism to ensure the integrity of the propagated version number as in [5]. Consequently, if a malicious node transmits a DIO message containing an incremented version number, the ensuing consequence is the initiation of the global repair procedure is initiated. This can precipitate topological inconsistencies and routing loops, particularly when the malicious node resides at a far distance from the root node. Moreover, among the RPL attacks, the version number attack significantly impacted the low-power wide-area network (LLN) [5].

1.2.3 Decrease rank attack (DR)

In this attack, the malicious node changes the value of its rank to be lower than other nodes, hence attracting other nodes to choose the opponent as their preferred parent.

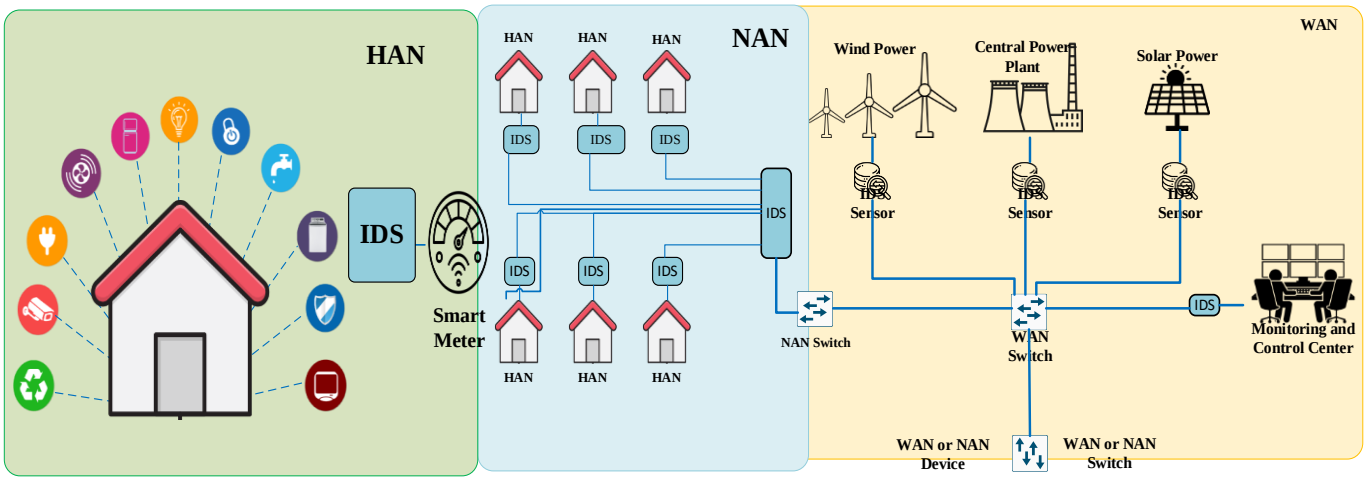

Figure 1. IDS systems framework in a smart grid network

1.3 Other work of attack on RPL

The RPL network is susceptible to various attacks, and a subset of these attacks are specific to RPL, such as the version number attack. Other attacks are inherited from wireless sensor network (WSN) networks, such as sinkhole attacks. Such attacks often can consume the resources of the RPL network and jeopardize the security triad, which includes confidentiality, integrity, and availability [12]. As a result, RPL is less encouraged to be deployed in IoT in sensitive applications. This section shows an overview of relevant existing studies in the field of IoT-based RPL networks in the literature.

The first IDS system to protect the RPL network was proposed by Raza et al. [13] and named (SVELTE), which means elegantly slim, this is because it considers the resource-constrained nature of IoT devices and has low overhead. Their IDS is a hybrid that uses signature and anomaly methods to detect two attacks, namely sinkhole and selective forwarding attacks. Moreover, the placement also uses a hybrid approach, with the IDS models Placed in the root and constrained nodes. The experiments were conducted on small to medium networks (i.e., 8,16, and 32 nodes with one, two, and four malicious nodes, respectively), based on Contiki OS. SVELTE includes three main models placed in the root node: 6Mapper, IDS, and mini firewall. SVELTE achieved success, especially in terms of the packet delivery rate and true positive rate. However, it has shortcomings, such as it detects attacks inherited from wireless sensor networks but does not focus on RPL-specific attacks. Moreover, it suffers from high false alarms and requires space to store attack signatures, and Mapper is subject to single failure for the IDS [7].

The core contribution of this study is summarized as follows:

·Propose a network IDS based on ensemble learning methods to shield the PRL network from three RPL attacks, both attacks that exploit the functionality of RPL and attacks inherited from wireless sensor networks. Moreover:

·Using a distributed sniffer device that intercepts the packets and sends them to an external server, without burdening the RPL network with additional overhead communication for an already constrained network.

·Identifies and highlights the most crucial features that are important for detecting the three mentioned attacks.

·Highlight the best performance algorithms used for each attack.

·Using K-fold validation to enhance the performance of the model.

This study is structured as follows: Section 1 provides background information and an introduction as well as discusses the attacks that may occur in the RPL smart grid network, along with related previous works. Section 2 details the methodology, including data analysis, dataset description, preprocessing steps, and the machine learning algorithms used. Section 3 presents the implementation results, evaluating the model using different metrics such as accuracy, precision, and F1 score, and summarizes the results of all the algorithms. Section 4 discusses these results. Finally, Section 5 presents the conclusions at the end.

2. METHODOLOGY

This section outlines the specific process involved in building the model, detailing the algorithm workflow, model training, and evaluation metrics.

2.1 IRAD dataset feature

The IRAD dataset [7] is a synthetic dataset generated using the COOJA simulation tool [14], which is part of Contiki OS [15]. COOJA allows for the simulation of various network environments, including the IoT, Mobile Ad Hoc Networks (MANET), and vehicle ad hoc network (VANET). In this study several attack scenarios were developed a cross different IoT nodes, with node size ranging from 10 to 1000 and varying percentage of malicious nodes (e.g., 5%, 10%, 20%). Feature extraction was applied to the dataset, resulting in a total of 18 features, as detailed in Table 1.

Table 1. The features of the IRAD dataset

Feature Number	Feature/ Abbreviations	Description
0	No.	Packet seq. nr.
1	Time	Simulation time
2	Source	Source node IP
3	Destination	Destination node IP
4	Info	Packet length
5	Info	Packet information
6	TR	Transmission rate
7	RR	Reception rate
8	TAT	Transmission avg. time
9	RAT	Reception avg. time
10	TPC	Transmitted packets
11	RPC	Received packets
12	TTT	Total transmission time
13	TRT	Total reception time
14	DAO	DAO packets
15	DIS	DIS packets
16	DIO	DIO packets
17	Label	Benign/malicious

2.2 Dataset preprocessing

The initial step in developing the IDS involves collecting sufficient network data traffic. These data should encompass both normal network behavior and abnormal states generated by various types of attacks. Packet sniffers are employed to collect this data ensuring the appropriate network attributes (or features) are captured for IDS development. Protecting RPL from external attacks, where external networks are integrated and exposed to common network threats, necessitates the collection of data with multiple network attributes to create an efficient IDS capable of identifying diverse cyber-attacks. Key network attributes include packet length, data transmission rate, throughput, time of inter-arrival, segment size, and active/idle duration, among others.

Due to the high degree of data dimensionality, preprocessing is essential to make the data more suitable for the IDS building process. The preprocessing steps includes:

2.2.1 Normalization

Machine learning algorithms perform better with normalized data, by converting each numerical value to a range of 0 to 1, as shown:

$$\text{Normalization} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

2.2.2 One-hot encoding

One-Hot encoding is applied to classify features to enhance

the efficiency of machine learning training by reducing the variance and skewness of the data distribution.

2.2.3 SMOTE

Typically, attack scenarios are less frequent to occur than normal behaviors, leading to a class imbalance, which in turn results in a class imbalance, which in turn results in a low anomaly detection rate. To address this, the synthetic minority over-sampling technique SMOTE is implemented [16]. SMOTE adds additional data for minority classes that lack sufficient representation, thus increasing the overall anomaly detection capacity.

2.3 Feature extraction

A total of 18 features were extracted from the dataset, which includes packet length, data transmission rate, throughput, inter-arrival segment size, and active/idle duration. Feature extraction reduces the dimensionality of the data and enhances the accuracy and effectiveness of the IDS. These features were vital for accurately detecting various types of cyber-attacks in the network.

2.4 Machine learning algorithms

In this article, we build an IDS using Tree-based algorithms using an ensemble learning technique model to detect various types of routing attacks. There are two hypotheses of attacks: (i) $\mathcal{H}_0$ when there are no attacks and (ii) $\mathcal{H}_1$ when there is at least one attack as shown:

$$\mathcal{H}_0: y[n] = \sum_j \sqrt{Ph\,\|x\|} + z[n] \quad (2)$$

$$\mathcal{H}_1: y[n] = \sum_j \sqrt{Ph\|x\|} + \sum_i \sqrt{Ph\,\|x\|}\, s[n] + z[n] \quad (3)$$

where, $y[n]$ is the received signal at time n, and $P\,h||x||$ is the power of the channel gain and the transmit signal of signal j, respectively. $s[n]$ is the sniffer signal for the attack signal x of at least 1 attacker.

We selected three tree-based to build the model: Decision Tree (DT), Random Forest (RF), Extra Trees (ET). These algorithms were chosen due to their advantages in classification results, such as improving the prediction accuracy and robust model sturdiness. Moreover, ensemble learning combines several models to classify one problem; hence, the prediction accuracy is higher.

2.4.1 Decision tree (DT)

A decision tree is a machine-learning algorithm used for classification and regression. It builds a flow chart-like tree where the nodes represent the features and branches describe the rules as the leaf node denotes the results [17].

2.4.2 Random forest (RF)

One of the machine learning algorithms used for both classification and regression uses an ensemble classifier where the outcome depends on their majority vote rules. Leading to the high class as the final result [18].

2.4.3 Extra trees (ET)

Another example of an ensemble model is extra trees (ET), which is a collection of random decision trees that are generated by processing various subsets of a data set [18].

2.5 Ensemble learning technique

A stacking ensemble method was employed, combining multiple base learner (DT, RF, ET) to produce a meta-learner that can make better predictions than single base learner alone. The base learners are trained on different subsets of data, and their predictions are passed to the meta learner.

2.6 K-fold cross-validation stacking ensemble (ES) proposed

To enhance model performance, a proposed K-fold cross-validation stacking ensemble model is implemented. The following steps are undertaken:

1. Implementation of Cross-Validation for Base Learners:

Cross-validation is applied to each of the base models (DT, RF, ET) individually. Every model is trained on K-1 folds and validated on the left-out fold. This procedure is repeated such that each fold has a chance to serve on the test set.

2. Creation of Meta-Feature Set:

The data is divided into K-fold, and then in each fold, we train the based model using the training data and obtain the predictions on the validation fold. The result of these predictions is stored for future use. Create a new training dataset for the meta-learner by using the predictions from the base models as input features and the outcomes considered as target variables.

3. Training the Meta-Learner:

Cross-validation for the new dataset obtained (meta-features set) will used to train the meta-learner to enhance the based model results. To make sure robust performance we implement cross-validation for the meta-learner. We use meta-features from the left outfold of each base model to train the meta-learner and then validate the meta-learner on the same left-out fold used for generating the meta-features.

4. Evaluation:

Lastly, the evaluation of the performance of the stacked was done using metrics such as accuracy, precision, recall, and F1-score.

2.7 Algorithm workflow

The flow diagram for the system is depicted in Figure 2.

The process includes:

1. Importing necessary libraries and the IRAD dataset.
2. Splitting the dataset into training and testing sets.
3. Training the classifiers using the training datasets.
4. Predicting the labels of the testing dataset.
5. Calculating the performance metrics (accuracy, precision, recall, and F1-score).
6. The calculated metrics.

2.8 Hyperparameter tuning

Hyperparameter tuning techniques such as grid search or random search are leveraged to optimize the performance of the algorithms. These techniques assist finding the best set of Hyperparameters to enhance the model's accuracy and efficiency.

Table 2 presents the features of the IRAD dataset, which is critical in identify and classifying the different types of cyber-attacks in the network. Each attack type relies on different features:

Version number attack (VN): Uses features like

simulation time (F1), destination node IP (F3), and transmission rate (F6).

Hello flood attack (HF): Utilizes features such as simulation time (F1), destination node IP (F3), and packet sequence number (F0).

Decrease rank attack (DR): Relies on simulation time (F1), DIO packets (F16), and reception rate (F7).

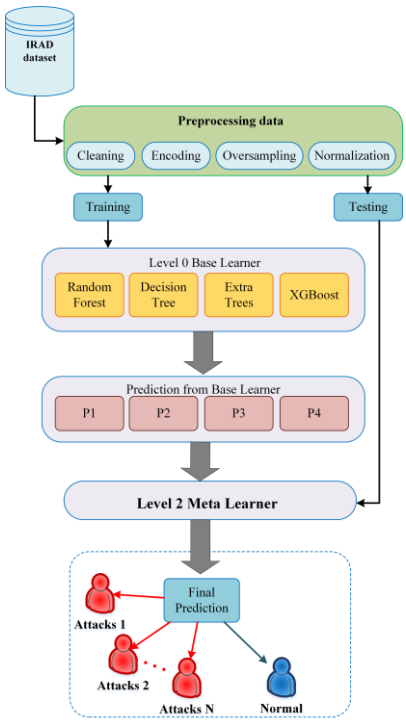

Figure 2. Workflow design of IDS

Table 2. The features of the IRAD dataset

Attack Type	Feature Number	Features	Description
Version	F1	TIME	Simulation time
Version	F3	Destination	Destination node IP
Version	F6	TR	Transmission rate
Hello flood	F3	Destination	Destination node IP
Hello flood	F1	Time	Simulation time
Hello flood	F0	No.	Packet seq. nr
Decrease rank	F1	Time	Simulation time
Decrease rank	F16	DIO	DIO packets
Decrease rank	F7	RR	Reception rate

3. IMPLEMENTATION RESULTS

3.1 Evaluation metrics

The system was executed by Python 3.5, installed on a laptop equipped with 6 Core i7-8700 CPU and 8 GB of RAM. We conducted a thorough analysis of our model, testing it under a variety of attack scenarios, using a single attack in conjunction with a legitimate dataset. This process was repeated for each of the three attacks, rank, hello flood, and version number. Our findings, reveal the results of the three attacks summarized in Tables 3-7 and Figures 3-7, and finally, Figure 8, which represents the comparative results with other models using the p-value.

We evaluate the model's performance by calculating the important metrics of accuracy, precision, recall, and F1 score by applying these equations.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{5}$$

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

$$F1 = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} = \frac{2\,TP}{2\,TP+FP+FN} \tag{7}$$

3.2 K-fold number of fold evaluation

K-fold evaluation is one type of cross-validation, which is the critical evaluation performance of machine learning.

We evaluated the number of folds to obtain the best results so we can use that in our proposed mode here. The K-fold number of 5 provides the highest accuracy precision and recall.

Table 3 and Figure 3 show that using 5-fold cross-validation yields the best results concerning accuracy, precision, and recall. This meant to divide the dataset into five equal-size subsets. During each iteration, four subsets are going to be utilized for training purposes, and the remaining subset will be used for validation. This process will be repeated five times with each subset serving as the validation set once.

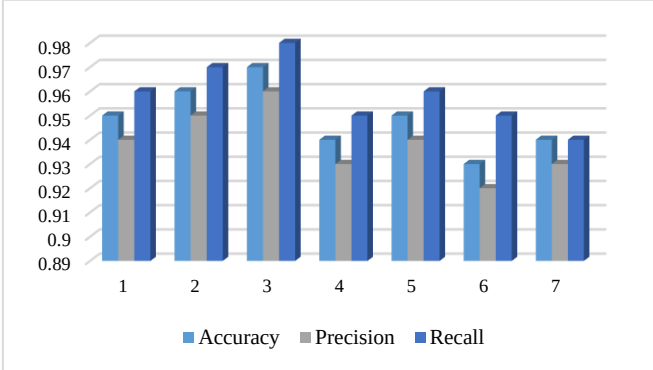

Figure 3. Stacking ensemble using 5-fold cross-validation

Table 3. Performance of stacking ensemble using 5-fold cross-validation

Fold Number	Accuracy	Precision	Recall
1	0.95	0.94	0.96
2	0.96	0.95	0.97
3	0.97	0.96	0.98
4	0.94	0.93	0.95
5	0.95	0.94	0.96
6	0.93	0.92	0.95
7	0.94	0.93	0.94
Average	**0.954**	**0.944**	**0.964**

3.3 Results of all algorithms

It is clear from the results that our model outperforms its predecessor across all three attack scenarios. When examining

the decrease rank, hello flood, and version number attacks, it was found that our model achieved an impressive approximately 0.999 accuracy, precision, recall, and F1 score which were obtained through the stacking method, reflecting a considerable improvement over the prior models. as we explore the ensemble algorithms on which our model is based.

In Table 4 and Figure 4, the results of the VN attack show that RF exhibited the highest performance among the algorithms, achieving an accuracy of 0.9998. This indicates that RF has superior capability in detecting VN attacks.

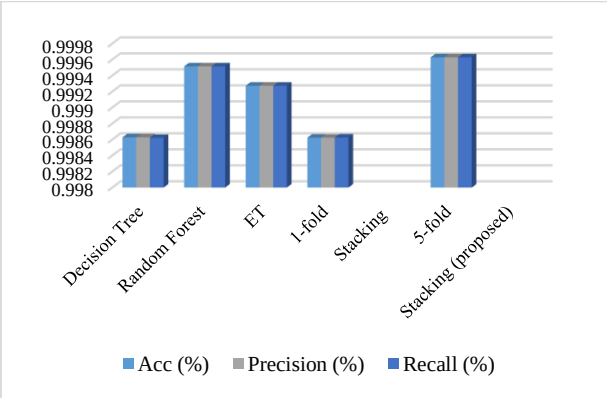

Figure 4. Version number (VN) attack results

Table 4. Version number (VN) attack results

Method	Acc (%)	Precision (%)	Recall (%)	F1 Score (%)
Decision tree	0.998628	0.998628	0.99862	0.9986
Random forest	0.999515	0.999515	0.999515	0.99951
ET	0.999273	0.999273	0.999273	0.99927
1-fold stacking	0.998624	0.998623	0.998625	0.99832
5-fold stacking (proposed)	0.999628	0.999628	0.999628	0.99962

Similarly, in the decreased rank attack Table 5 and Figure 5, the RF algorithm outperforms the other algorithms in terms of efficiency.

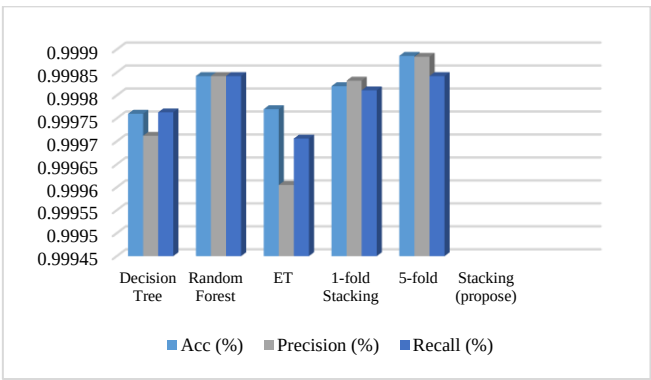

Figure 5. Decrease rank (DR) attack results

Contrastingly, according to Table 6 and Figure 6, the ET algorithm shows excellent performance by achieving perfect scores is that 1.0 in all metrics. Hence it is correctly detecting all HF attacks and zero false predictions. In contrast, other algorithms obtained lower scores. Stacking has a score of 0.9987 which indicates significant prediction.

Table 7 and Figure 7 presents the results of our model conducted on the entire dataset by merging the three RPL attacks and the legitimate dataset into one dataset (mixed data). The results obtained based on this dataset are promising, achieving a proximity of 0.998 in all metrics, as demonstrated in Table 6.

To evaluate our model, we compare it with related work that used the same attacks or the same dataset. Table 8 presents the comparative results between the proposed model and others, achieving superior performance across the board in terms of accuracy, precision, recall, and F1 score. Indicating that it is the most effective and robust model for detecting RPL network attacks.

Table 5. Decrease rank **(**DR) attack results

Method	Acc (%)	Precision (%)	Recall (%)	F1 Score (%)
Decision tree	0.999760	0.999712	0.999763	0.999763
Random forest	0.999842	0.999842	0.999842	0.999842
ET	0.999770	0.999605	0.999706	0.999241
1-fold stacking	0.999820	0.999832	0.999811	0.999812
5-fold stacking (propose)	0.999886	0.999884	0.999842	0.99982

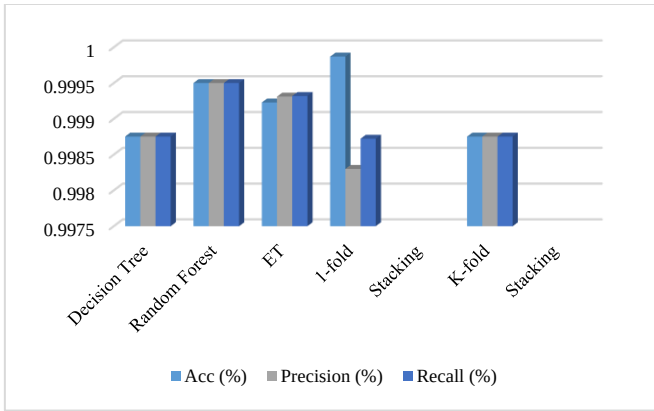

Figure 6. Hello-flood (HF) attack results

Table 6. Hello-flood (HF) attack results

Method	Acc (%)	Precision (%)	Recall (%)	F1 Score (%)
Decision tree	0.998753	0.998753	0.998753	0.998753
Random forest	0.999501	0.999501	0.999501	0.999501
ET	0.999230	0.999312	0.999320	0.999311
1-fold stacking	0.999871	0.998301	0.998722	0.998501
K-fold stacking	0.998753	0.998753	0.998753	0.998753

Lastly, the bar chart presented in Figure 8 shows the relative results with p-values across multiple metrics. All of the approaches are more accurate than 97%, but the Proposed 5-fold strategy performs the best, averaging around 99.69%. Precision ranges from roughly 97% to almost 100%, with the

Proposed 5-fold technique once more delivering the best results. While recall scores differ, the proposed 5-fold is almost 100%. Similar trends can be seen in the F1 Score, where the Proposed technique performs better than the others. P-values in both cases show statistical significance. In this dataset, the Proposed 5-fold approach performs outstandingly overall across all variables that were measured.

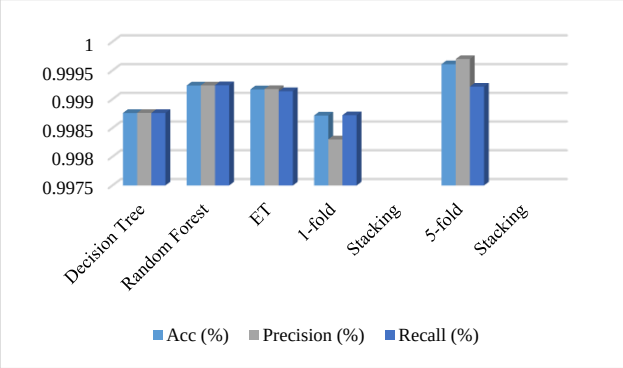

Figure 7. Model for all attacks

Table 7. Evaluation performance summary for our model on all attacks together

Method	Acc (%)	Precision (%)	Recall (%)	F1 Score (%)
Decision tree	0.998761	0.998762	0.998761	0.99876
Random forest	0.999240	0.999241	0.999244	0.999240
ET	0.999172	0.999178	0.999141	0.999154
1-fold stacking	0.998717	0.998302	0.998722	0.998505
5-fold stacking	0.99961	0.99970	0.99922	0.99970

4. DISCUSSION

With a particular focus on three well-known attack scenarios—rank, hello flood, and version number attacks—our study assessed the effectiveness of a novel model intended to identify and counteract network attacks. Using Python 3.5 on a PC with an Intel Core i7-8700 CPU and 8 GB RAM, we carried out a number of experiments to evaluate the effectiveness of the model in different simulated scenarios. First off, our model performed admirably on all assessed measures, including F1 score, accuracy, precision, and recall. Compared to individual algorithms, the application of ensemble approaches, especially stacking, greatly increased the capabilities of our model. Table 3, for example, shows that using 5-fold cross-validation produced average results of 95.4% accuracy, 94.4% precision, and 96.4% recall, which is a continuous improvement over single-fold assessment. When comparing our findings to the body of literature currently in publication, our model continuously beat earlier methods in terms of robustness and accuracy. Table 8 shows that when compared to previous ensemble approaches and individual algorithms employed in similar research, our suggested model performed better across all measures (accuracy, precision, recall, and F1 score). These results highlight how well our strategy works to strengthen security defenses against network intrusions.

The efficacy of group techniques, as demonstrated by random forest's detection of VN attacks (Table 4), implies that merging different algorithms can lessen the weaknesses present in single strategies. Our research shows how sophisticated machine learning methods can strengthen defenses against changing network threats, which has important applications for cybersecurity. Even though our model shows good accuracy and precision, real-time deployment and scalability problems with bigger datasets are yet unresolved.

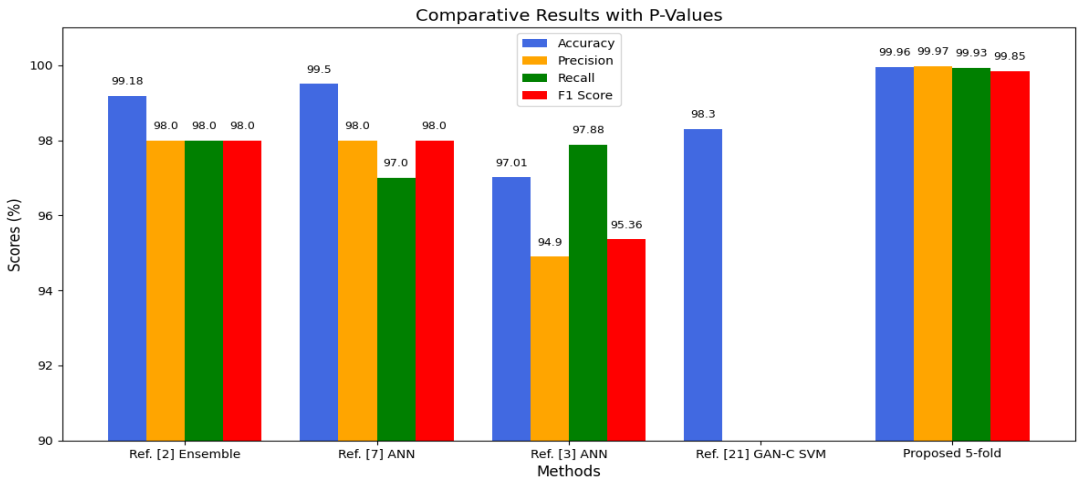

Figure 8. Comparative results with bar virtualization

Table 8. Comparative results summary between the proposed model and others

Reference	Methods Used	Acc (%)	Precision (%)	Recall (%)	F1 Score (%)	Attacks	Dataset Used
Ref. [7]	ANN	0.995	0.98	0.97	0.98	DR, HF, VN	IRAD
Ref. [19]	Ensemble learning	0.9918	0.98	0.98	0.98	DR, HF, VN	Author generated
Ref. [20]	GAN-C, SVM	0.983	-	-	-	DR, HF, VN	IRAD
Ref. [21]	ANN	0.9701	0.949	0.9788	0.9536	DR, HF, VN	IRAD
Proposed model 5-fold	Ensemble learning	0.99961	0.999753	0.9993	0.99853	DR, HF, VN	IRAD

5. CONCLUSIONS

In summary, our study tackles the crucial problem of protecting IoT networks from a variety of complex threats by creating and assessing a cutting-edge IDS. By utilizing ensemble learning methods, namely a 5-fold cross-validated strategy, our IDS exhibits remarkable efficacy in identifying and counteracting assaults directed on RPL networks, such as VN, HF, and DR. The key innovation of our study is the use of ensemble learning, which combines the advantages of several methods to improve detection resilience and accuracy. Our model outperformed other models in the literature with an amazing accuracy of 99.88% along with good precision, recall, and F1 scores. Comparative results and statistically significant p-values highlight this progress and confirm superiority of our approach. Additionally, by offering a scalable and practical method for improving the security posture of IoT networks, our research advances the field. Our IDS provides an excellent means of detecting and countering routing attacks, making it a useful tool for addressing cyber threats in real-time situations. Future research can further enhance IoT security frameworks by exploring additional ensemble methodologies, scaling the model for bigger networks, and incorporating real-time threat intelligence.

ACKNOWLEDGMENT

This work is supported by the Indian Council for Cultural Relations ICCR. And it is also supported by University of Mosul.

REFERENCES

[1] Muralidharan, C., Mohamed Sirajudeen, Y., Anitha, R. (2021). Synergy of Internet of Things with cloud, artificial intelligence and blockchain for empowering autonomous vehicles. In: Ahmed, K.R., Hassanien, A.E. (eds) Deep Learning and Big Data for Intelligent Transportation. Studies in Computational Intelligence, vol 945. Springer, Cham, pp. 225-244. https://doi.org/10.1007/978-3-030-65661-4_11

[2] Al-Amiedy, T.A., Anbar, M., Belaton, B., Bahashwan, A.A., Hasbullah, I.H., Aladaileh, M.A., Mukhaini, G.A. (2023). A systematic literature review on attacks defense mechanisms in RPL-based 6LoWPAN of Internet of Things. Internet of Things, 22: 100741. https://doi.org/10.1016/j.iot.2023.100741

[3] Lahbib, A., Toumi, K., Elleuch, S., Laouiti, A., Martin, S. (2017). Link reliable and trust aware RPL routing protocol for Internet of Things. In 2017 IEEE 16th International Symposium on Network Computing and Applications (NCA), Cambridge, MA, USA, pp. 1-5. https://doi.org/10.1109/NCA.2017.8171360

[4] Wadhaj, I., Ghaleb, B., Thomson, C., Al-Dubai, A., Buchanan, W.J. (2020). Mitigation mechanisms against the DAO attack on the routing protocol for low power and lossy networks (RPL). IEEE Access, 8: 43665-43675. https://doi.org/10.1109/ACCESS.2020.2977476

[5] Almusaylim, Z.A., Alhumam, A., Jhanjhi, N.Z. (2020). Proposing a secure RPL based Internet of Things routing protocol: A review. Ad Hoc Networks, 101: 102096. https://doi.org/10.1016/j.adhoc.2020.102096

[6] Swessi, D., Idoudi, H. (2022). A survey on Internet-of-Things security: Threats and emerging countermeasures. Wireless Personal Communications, 124(2): 1557-1592. https://doi.org/10.1007/s11277-021-09420-0

[7] Yavuz, F.Y., Unal, D., Gül, E. (2018). Deep learning for detection of routing attacks in the Internet of Things. International Journal of Computational Intelligence Systems, 12(1): 39-58. https://doi.org/10.2991/ijcis.2018.25905181

[8] Winter, T., Thubert, P., Brandt, A., Hui, J., Kelsey, R., Levis, P., Pister, K., Struik, R., Vasseur, J.P., Alexander, R. (2012). RPL: IPv6 routing protocol for low-power and lossy networks (No. RFC 6550). Internet Engineering Task Force (IETF), Wilmington, DE, USA. https://www.rfc-editor.org/rfc/rfc6550.html.

[9] Mayzaud, A., Badonnel, R., Chrisment, I. (2016). A taxonomy of attacks in RPL-based Internet of Things. International Journal of Network Security, 18(3): 459-473. https://doi.org/10.6633/IJNS.201605.18(3).07

[10] Canbalaban, E., Sen, S. (2020). A cross-layer intrusion detection system for RPL-based Internet of Things. In Proceedings of Ad-Hoc, Mobile, and Wireless Networks: 19th International Conference on Ad-Hoc Networks and Wireless (ADHOCNOW 2020), Bari, Italy, pp. 214-227. https://doi.org/10.1007/978-3-030-61746-2_16

[11] Le, A., Loo, J., Luo, Y., Lasebae, A. (2013). The impacts of internal threats towards routing protocol for low power and lossy network performance. In 2013 IEEE Symposium on Computers and Communications (ISCC), Split, Croatia, pp. 789-794. https://doi.org/10.1109/ISCC.2013.6755045

[12] Pasikhani, A.M., Clark, J.A., Gope, P., Alshahrani, A. (2021). Intrusion detection systems in RPL-based 6LoWPAN: A systematic literature review. IEEE Sensors Journal, 21(11): 12940-12968. https://doi.org/10.1109/JSEN.2021.3068240

[13] Raza, S., Wallgren, L., Voigt, T. (2013). Svelte: Real-time intrusion detection in the Internet of Things. Ad Hoc Networks, 11(8): 2661-2674. https://doi.org/10.1016/j.adhoc.2013.04.014

[14] Osterlind, F., Dunkels, A., Eriksson, J., Finne, N., Voigt, T. (2006). Cross-level sensor network simulation with COOJA. In Proceedings of 2006 31st IEEE Conference on Local Computer Networks, Tampa, FL, USA, pp. 641-648. https://doi.org/10.1109/LCN.2006.322172

[15] Contiki: The open-source operating system for the Internet of Things. http://www.contiki-os.org/.

[16] Arjunan, K., Modi, C.N. (2017). An enhanced intrusion detection framework for securing the network layer of cloud computing. In 2017 ISEA Asia Security and Privacy (ISEASP), Surat, India, pp. 1-10. https://doi.org/10.1109/ISEASP.2017.7976988

[17] Wei, G., Mu, W., Song, Y., Dou, J. (2022). An improved and random synthetic minority oversampling technique for imbalanced data. Knowledge-Based Systems, 248: 108839. https://doi.org/10.1016/j.knosys.2022.108839

[18] Parmar, A., Katariya, R., Patel, V. (2019). A review on random forest: An ensemble classifier. In: Hemanth, J., Fernando, X., Lafata, P., Baig, Z. (eds) International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018. Lecture Notes on Data Engineering and Communications Technologies, vol 26. Springer, Cham, pp. 758-763. https://doi.org/10.1007/978-3-030-03146-6_86

[19] Verma, A., Ranga, V. (2019). ELNIDS: Ensemble learning based network intrusion detection system for RPL-based Internet of Things. In 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, pp. 1-6. https://doi.org/10.1109/IoT-SIU.2019.8777504

[20] Nayak, S., Ahmed, N., Misra, S. (2022). Deep learning-based reliable routing attack detection mechanism for industrial Internet of Things. Ad Hoc Networks, 123: 102661. https://doi.org/10.1016/j.adhoc.2021.102661

[21] Osman, M., He, J., Zhu, N., Mokbal, F.M.M. (2024). An ensemble learning framework for the detection of RPL attacks in IoT networks based on the genetic feature selection approach. Ad Hoc Networks, 152: 103331. https://doi.org/10.1016/j.adhoc.2023.103331