Enhancing IoT Network Security: Machine Learning Solutions for Securing RPL-Based IoT Networks from Routing Attacks

Kabil Pranav K, Niranjan K, V Krishna Moorthy, Vishveshwaran K, Gandhiraj R*
Department of Electronics and Communication Engineering
Amrita School of Engineering, Coimbatore,
Amrita Vishwa Vidyapeetham, India
r_gandhiraj@cb.amrita.edu*

***Abstract*—Routing attacks are the major threat to RPL based networks so this research aims to develop a machine learning model which will identify various routing attacks and even the new form of routing attacks in the RPL networks. Building a machine learning model capable of identifying routing attack threats precisely in networks with RPL support. Using a network traffic dataset to assess the efficacy of the ML models. Showcasing the model's resistance to various forms of assault. The project's outcomes reveal that the suggested machine learning model has great potential of identifying routing attacks, since it is highly accurate and resistant to emerging attacks. It improves the overall security of the RPL networks as a result.**

***Index Terms*—Routing Attack, RPL-Iot Network, Machine Learning, NetSim.**

I. Introduction

Internet of Things [1] uses 6LOWPAN protocol to connect with different things. This protocol use less power consumption and works with low power network things. It operates at 20-250 kbps for a smart network with Destination Oriented Directed Acyclic Graph (DODAG) to transfer data efficiently. In a DODAG the control packets are calculated by "Rank " elements to make robust and scalable network to improve overall performance[2].

"The ML based approach to detect Routing attack in RPL Networks" study by Raveendranadh Bokka[3]. The dataset from the paper consist of RPL_RANK, DIO_count , PACKET_COUNT, and labels to detect Normal and Attack data. The study paper consisted various types of attack like sinkhole, blackhole, selective Forwarding, sybil, DIO flooding and more. They used various ML models like Decision Trees, Naive Bayes, AdaBoost, K-Neighbours, Logistic Regression, and they used area under the ROC curve to validate the system's(models) performance[4].

RPL based routing protocol faces issues like security, sinkhole or blackhole attacks. RPL networks does not detect important or valid messages. Hence they are prone to attacks and are vulnerable to networks[5].

In order to handle the network safely we need a efficient and smart technique. Our aim is to handle security based issues in a RPL_based IOT networks[6]. A ML based model is implemented to detect attacks and run efficiently. Our research paper deals with blackhole and sinkhole attacks and improve the efficiency of the system . The network is implemented in a real life situation scenario. By this RPL based IOT networks are more secure[3].

II. Methodology

This section outlines the approach we used to identify routing attacks in RPL based Iot networks through our research. We create a malicious node and created dataset NetSim environment and add machine learning classifiers to it.

A. Dataset on RPL Attacks

The NetSim software application was used to construct the synthetic RPL assaults dataset, which was used for the research. Machine learning classifiers are trained and assessed using this dataset. There are 21 characteristics in all.

TABLE I
Dataset Features

PACKET ID
PACKET TYPE
CONTROL PACKET TYPE/APP
SOURCE ID
DESTINATION ID
TRANSMITTER ID
RECEIVER ID
NETWORK LAYER ARRIVAL TIME
MAC LAYER ARRIVAL TIME
PHYSICAL LAYER ARRIVAL TIME
PHYSICAL LAYER START TIME
PHYSICAL LAYER END TIME
NETWORK LAYER PAYLOAD
MAC LAYER PAYLOAD
PHYSICAL LAYER PAYLOAD
PHYSICAL LAYER OVERHEAD
PACKET STATUS
SOURCE IP
DESTINATION IP
GATEWAY IP
NEXT HOP IP

And these feature were selected based on the analysing the dataset using the correlation matrix below:

It also includes two labeling attributes, "Successful" and "Collided". The dataset comprises a range of routing attack types, including Sinkhole and Blackhole. The dataset description is presented in Table 1.

B. Implementation Steps

We followed a systematic approach. It includes feature extraction, data pre - processing (includes removal of none values) and the application of various machine learning algorithms. Out of all models that were implemented, Random

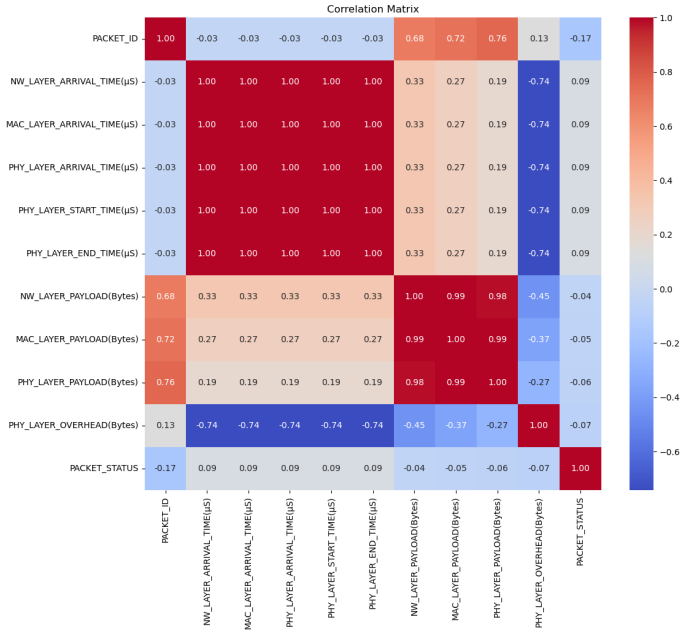

Fig. 1. Correlation Matrix

TABLE II
DATASET DETAILS

Classification	Number of Instances
Malicious	21,527
Non-Malicious	49,265

Forest proven to be the most efficient model to reduce the dataset dimensions by selecting the most relevant features which gives higher accuracy [7][8].

1) Data Preprocessing: Involves several essential steps as follows: First by removing missing values, encoding categorical data into numeric using the Label Encoder, and assigning Attack as 1 and Normal as 0.

2) Data Separation: As training and testing data are necessary for the evaluation of machine learning models, We conducted evaluations of these machine learning models with different train-test dataset size from the RPL dataset that were generated using netsim,training percentage like - 60%, 70%, 80%, and 90% and remaining for testing[3].

3) Machine Learning Algorithms: The performance of different machine learning algorithms, each giving unique characteristics and advantages [9]:

I. **K-Nearest Neighbors (KNN):** KNN focuses in identifying data points that is similar to existing ones within the same dataset. It is particularly suited for multidimensional data which is especially efficient during the training phase and slightly slower during estimation.

II. **Gaussian Naive Bayes (GNB):**Gaussian naïve bayes is a machine learning algorithm which will categories data based on features. Classification is done based on the gaussian distribution. This model first assumes the data are independent to each other so during training this model can learn mean and standard deviation for each category. While prediction, the data are classified and picks the one which is more suitable.

III. **Random Forest (RF):** It consists of multiple decision trees. They are combined together to get more accurate predictions. Out of all models, it is more popular because of its flexilbility in handling classification and regression problems. The model tends to give higher accuracy and can overcome the problem of overfitting as the number of trees gets increased.[10].

IV. **Decision Tree Classifier (DT):** This classifier specializes in dividing datasets into distinct groups. It takes two inputs: a data array, with dimensions of m samples and n features, containing training samples, and class labels for the training data. In the event that multiple classes share the same highest probability, the classifier predicts the class with the lowest index. Instead of producing a specific class, the Decision Tree (DT) predicts the similarity in each class, represent the amount of training samples that are present within the leaf.

V. **Support Vector Machine (SVM):** It is a machine learning algorithm designed for classification purpose. It finds the best hyperplane to separate different class. The decision boundary is defined from the data points that are most suitable for it. Kernel trick helps to transform the non-linear data into higher dimensions. This model maximizes the margin.

VI. **Adaptive Boosting (AdaBoost):** Adaboost is machine learning algorithm which is a combination of simple models to make one strong model. This model gives misclassified data points more importance because the next model will focus on what the previous model has got wrong. So this machine learning algorithm works on the iteration process which will correct the mistakes of the previous.

VII. **Multilayer Perceptron (MLP):**It is a neural network that has different layers of interconnected neurons. Neurons can be combined together to create a neural network.. MLPs encompass three levels: input, output, and hidden layers allowing for complicated supervised learning.

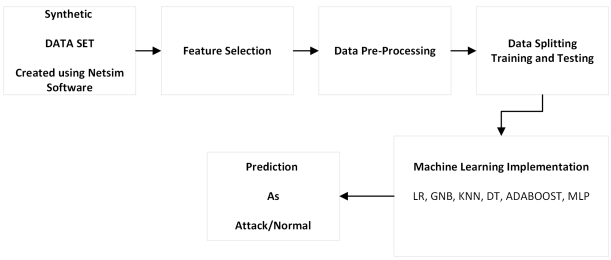

Fig. 2. Proposed Implementation Flowgraph

III. Results and Analysis

The analysis of this paper is a sinkhole attack and blackhole attack in RPL networks simulated in a NetSim Software. The network data is analysed and processed to create a dataset to train the ML model[11].

A. Simulation Setup

The network simulation is created using the following steps[5]:

- IOT network is created in Netsim consisting of nine wireless sensor node supporting RPL, a 6LoWPAN gateway and router.
- Data is collected by enabling the packet tracing and Wireshark to get data from all nodes.
- The simulation of the RPL based network is set to 1000 sec for DODAG(Destination Oriented Directed Acyclic Graph) to develop the network. The root node of DODAG is its gateway.

3.

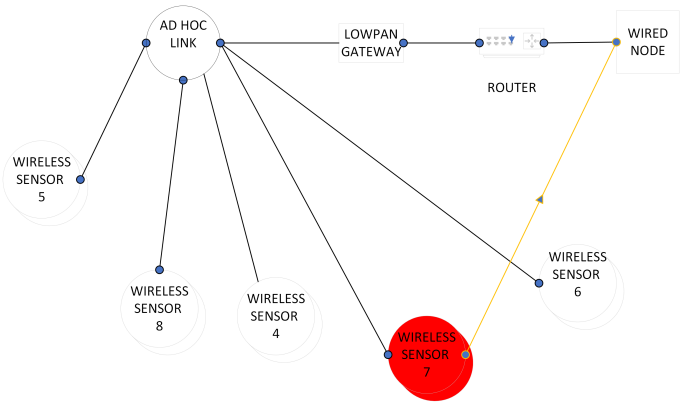

Fig. 3. Iot Network - Setup

B. Sinkhole Attack Introduction

The impact and working of sinkhole attack is implemented into the network by RPL rank[7][(5)].The rank is created in the following steps:

- The RPL rank is assigned for 2 nodes in the malicious,c file to make them as sinkhole node in the given Fig 4.
- These sinkhole nodes are addressed as fake rank to attract data traffic to itself.
- The packet trace is monitored observed that the packets are left at malicious node when routing among all nodes surrounding the malicious node .
- The throughput of the data is reduced to zero as packets reach sinkhole nodes and won't transfer or forward data from it.[5]

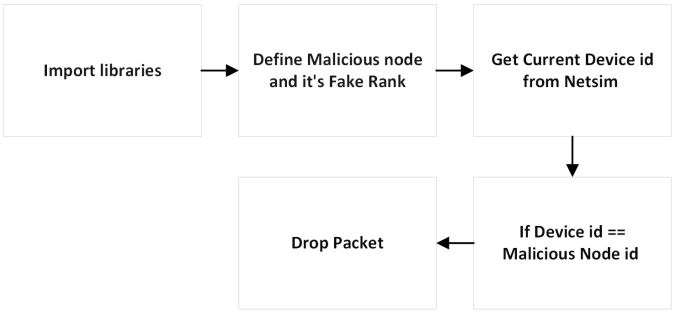

Fig. 4. Malicious Nodes in Operation

C. Analysis of Network Behavior

The analysis from the network performance is that parent selection of legitimate based modes are selected [12][2].The observations are :

- Before sinkholes are analyzed or detected the parent selection of nodes are chosen.
- The parent nodes are switched on fake lower rank by the sinkhole nodes create a significant impact on network structure. The following is visually visible in Figure 5.

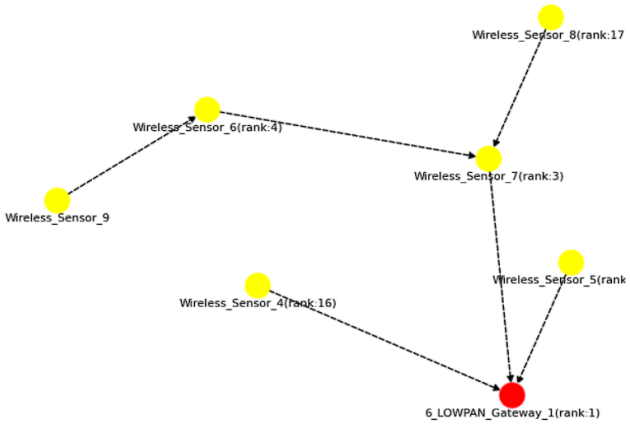

Fig. 5. DODAG Formation

D. Dataset Generation

According to the simulation the dataset is processed which includes throughput, packet losses, parent modification and other topics. The sinkhole node availability is also included to label the dataset. The aim of the ML model is to detect sinkhole attack [13].

A aforementioned statistics , were used as an essential tool for analysis of malicious nodes and establishment of DODAG network structure[6].

The dataset and outcomes are used to build a strong ML model to secure RPL based IOT networks from sinkhole attacks.

E. Machine Learning Model Implementation

After simulating in Netsim, the dataset has been generated and then the Netsim software applied machine learning models which identifies the sinkhole attack which enhances the network security. The key outcomes include::

- Random forest (RF) and Support Vector Machine (SVM) has implemented.
- SVM and RF models achieved 96% accuracy each which indicates high efficiency in identifying sinkhole attacks.
- Graphical results of each model's performance has been plotted.

1) SVM Model Results: The SVM model provides an accuracy of 96% which is an exceptional performance. The results are depicted in Figure 6.

SVM Classification Report:

	precision	recall	f1-score	support
0	0.92	0.91	0.92	19012
1	0.98	0.98	0.98	39405
accuracy			0.96	58417
macro avg	0.95	0.94	0.95	58417
weighted avg	0.96	0.95	0.96	58417

Fig. 6. SVM Model Results

2) Random Forest Model Results: Random Forest Model provides an accuracy of 96% which proves to be highly effective model. Figure provides a visual depiction of the findings 7.

Random Forest Classification Report:

	precision	recall	f1-score	support
0	0.97	0.95	0.97	19012
1	0.96	0.96	0.98	39405
accuracy			0.96	58417
macro avg	0.96	0.95	0.97	58417
weighted avg	0.96	0.95	0.97	58417

Fig. 7. Random Forest Model Results

and the respective feature importance plot for training the random forest is given below:

F. Comparison of Model Performance

For different training and testing dataset sizes, we performed a comparison study in order to fully evaluate the performance of machine learning models.

From this demonstration, we learnt that ML models plays important role in protecting RPL-based IoT networks from sinkhole attacks . The SVM proves to be a reliable solution thus being effective in practical situations.

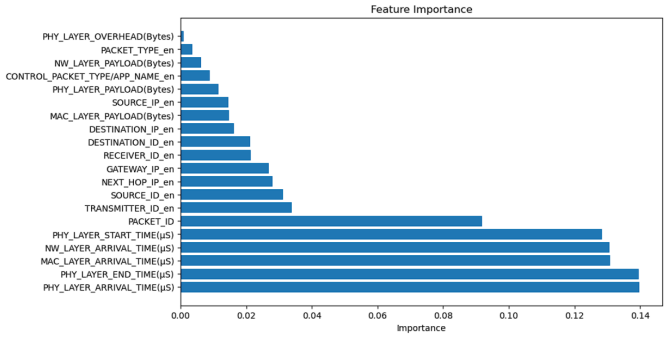

Fig. 8. Random Forest Feature Importance Plot

	Model	Training Set Size	Precision	Recall	F1 Score	Accuracy
0	Support Vector Machine	40	0.877358	0.935484	0.934673	0.919487
1	Support Vector Machine	30	0.867925	0.935484	0.929293	0.929133
2	Support Vector Machine	20	0.926415	0.917438	0.921990	0.962642
3	Support Vector Machine	10	0.905660	1.000000	0.950495	0.955660
4	Random Forest	40	0.913386	0.935484	0.924303	0.875566
5	Random Forest	30	0.912587	0.945652	0.928826	0.904214
6	Random Forest	20	0.974894	0.959730	0.972252	0.963208
7	Random Forest	10	0.947368	0.932144	0.942408	0.966226
8	Multi-Layer Perceptron	40	0.877358	1.000000	0.934673	0.817358
9	Multi-Layer Perceptron	30	0.867925	1.000000	0.929293	0.867925
10	Multi-Layer Perceptron	20	0.875598	0.989189	0.928934	0.897925
11	Multi-Layer Perceptron	10	0.933396	0.979167	0.910670	0.933396
12	Gaussian Naive Bayes	40	0.877358	1.000000	0.934673	0.877358
13	Gaussian Naive Bayes	30	0.867925	1.000000	0.929293	0.867925
14	Gaussian Naive Bayes	20	0.872642	1.000000	0.931990	0.872642
15	Gaussian Naive Bayes	10	0.905660	1.000000	0.950495	0.905660
16	K-Nearest Neighbors	40	0.902913	0.981183	0.925222	0.902913
17	K-Nearest Neighbors	30	0.910670	0.989130	0.923858	0.910670
18	K-Nearest Neighbors	20	0.935660	0.972973	0.918367	0.935660
19	K-Nearest Neighbors	10	0.923858	0.968750	0.934673	0.923858
20	Decision Tree	40	0.910290	0.927419	0.918775	0.856132
21	Decision Tree	30	0.914894	0.934783	0.924731	0.867925
22	Decision Tree	20	0.913514	0.913514	0.913514	0.849057
23	Decision Tree	10	0.945652	0.906250	0.925532	0.867925
24	AdaBoost	40	0.910670	0.986559	0.947097	0.903302
25	AdaBoost	30	0.898026	0.989130	0.941379	0.893082
26	AdaBoost	20	0.910000	0.983784	0.945455	0.900943
27	AdaBoost	10	0.940000	0.979167	0.959184	0.924528

Fig. 9. All Models Output Comparison

IV. Conclusion

RPL protocol has been widely used in IoT networks, but still it is vulnerable to routing attacks like sinkhole attack. Due to this malicious node advertise fake rank and decrease the network performance. Simulation results tells an idea about how the sinkhole attack affects the results of throughput and how the network is affected. Decision tree classifier machine learning algorithm can detect the RPL assault and with this it shows promises for creating the intrusion detection system

for RPL networks which will be effective and reliable as well. With assessment measures like precision, recall, and F1-score, the Decision tree classifier machine learning algorithm able to detect the RPL assaults over precision rate of 96%. In conclusion, simulation results are used to analyse the RPL networks assaults. Machine learning models does really helps in protecting the RPL-based IoT network. For the RPL protocol, a complete method for assessing vulnerabilities and creating smart security solutions combines network modelling and machine learning.

REFERENCES

[1] A. Paul and A. S. Pillai, "A review on rpl objective function improvements for iot applications," in *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*. IEEE, 2021, pp. 80–85.

[2] T. Winter, P. Thubert, A. Brandt, J. Hui, R. Kelsey, P. Levis, K. Pister, R. Struik, J.-P. Vasseur, and R. Alexander, "Rpl: Ipv6 routing protocol for low-power and lossy networks," Tech. Rep., 2012.

[3] R. Bokka and T. Sadasivam, "Machine learning techniques to detect routing attacks in rpl based internet of things networks."

[4] A. Sharma, H. Babbar, S. Rani, D. K. Sah, S. Sehar, and G. Gianini, "Mhseer: A meta-heuristic secure and energy-efficient routing protocol for wireless sensor network-based industrial iot," *Energies*, vol. 16, no. 10, p. 4198, 2023.

[5] J. Deogirikar and A. Vidhate, "Security attacks in iot: A survey," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE, 2017, pp. 32–37.

[6] M. Sharma, H. Elmiligi, F. Gebali, and A. Verma, "Simulating attacks for rpl and generating multi-class dataset for supervised machine learning," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2019, pp. 0020–0026.

[7] M. A. Kareem and S. Tayeb, "Ml-based nids to secure rpl from routing attacks," in *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2021, pp. 1000–1006.

[8] N. Panda and M. Supriya, "Blackhole attack impact analysis on low power lossy networks," in *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)*. IEEE, 2022, pp. 1–5.

[9] ——, "Blackhole attack prediction in wireless sensor networks using support vector machine," in *Advances in Signal Processing, Embedded Systems and IoT: Proceedings of Seventh ICMEET-2022*. Springer, 2023, pp. 321–331.

[10] R. B. L. S. S. V. V. Boobala Muralitharan D1, Priyanka P2, "Fake news detection using machine learning," *http://ijircce.com/admin/main/storage/app/pdf/1P7aXyh2UJuJg3Nfh5M6Jcw8CBClqvg5MyD0TXwt.pdf*, 2022.

[11] V. Tharunika, T. Shridhar, K. Veeresh, S. K. Thangavel, K. Srinivasan, S. Vajipayajula, and A. Tibrewal, "Detection and prevention of advanced persistent threat (apt) activities in heterogeneous networks using siem and deep learning," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2023, pp. 1–8.

[12] P. Shukla, "Ml-ids: A machine learning approach to detect wormhole attacks in internet of things," in *2017 Intelligent Systems Conference (IntelliSys)*. IEEE, 2017, pp. 234–240.

[13] I. Arnaldo, A. Cuesta-Infante, A. Arun, M. Lam, C. Bassias, and K. Veeramachaneni, "Learning representations for log data in cybersecurity," in *Cyber Security Cryptography and Machine Learning: First International Conference, CSCML 2017, Beer-Sheva, Israel, June 29-30, 2017, Proceedings 1*. Springer, 2017, pp. 250–268.