Customized dataset-based machine learning approach for black hole attack detection in mobile ad hoc networks

Houda Moudni[1], Mohamed Er-Rouidi[2], Mansour Lmkaiti[3], Hicham Mouncif[3]
[1]TIAD Laboratory, Faculty of Science and Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco
[2]Modeling and combinatorial laboratory, Department of Mathematics and Computer Science, Polydisciplinary Faculty, Cadi Ayyad University, Safi, Morocco
[3]LIMATI Laboratory, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco

Article Info

Article history:

Received Jul 30, 2024
Revised Sep 20, 2024
Accepted Oct 1, 2024

Keywords:

Black hole attack
Intrusion detection system
Machine learning
Mobile ad hoc networks
Security

ABSTRACT

This article explores the application of machine learning (ML) algorithms to classify the black hole attack in mobile ad hoc networks (MANETs). Black hole attacks threaten MANETs by disrupting communication and data transmission. The primary goal of this study is to develop an intrusion detection system (IDS) to detect and classify this attack. The research process involves feature selection, the creation of a custom dataset tailored to the characteristics of black hole attacks, and the evaluation of four machine learning models: random forest (RF), logistic regression (LR), k-nearest neighbors (k-NN), and decision tree (DT). The evaluation of these models demonstrates promising results, with significant improvements in accuracy, precision, F1-score, and recall metrics. The findings underscore the potential of machine learning in enhancing the security of MANETs by providing an effective means of attack classification.

Corresponding Author:

Houda Moudni
TIAD Laboratory, Faculty of Science and Technology, Sultan Moulay Slimane University
Av Med V, BP 591, Beni-Mellal 23000, Maroko
Email: h.moudni@usms.ma

1. INTRODUCTION

Mobile ad hoc networks (MANETs) [1] are self-configuring networks composed of mobile nodes that form a dynamic network infrastructure without relying on a centralized authority. MANETs have gained significant attention due to their flexibility in providing communication in various environments, making them ideal for emergency and military applications. However, the decentralized nature and dynamic topology of MANETs make them particularly vulnerable to a wide range of security threats, including routing attacks, denial of service, and black hole attacks [2]–[4].

Black hole attacks are particularly severe as they involve a malicious node falsely advertising the shortest path to the destination, subsequently intercepting and dropping all data packets passing through it. This not only disrupts network communication but also compromises data integrity and confidentiality. Traditional security mechanisms, such as encryption and authentication protocols, while effective in some scenarios, often fall short in addressing the sophisticated and adaptive nature of such attacks in MANETs. Existing solutions, including heuristic-based and rule-based intelligent intrusion detection systems (IDSs), are often limited by their inability to adapt to new and evolving attack patterns.

Given these constraints, there is an urgent need for intelligent systems that can dynamically detect and classify such attacks, enabling timely and effective countermeasures. Recent advances in machine learning (ML) [5]–[7] offer a promising approach to this challenge. ML algorithms can analyze large volumes of network data, identify patterns associated with attacks, and distinguish them from normal network behavior. By learning from past incidents, these algorithms can provide robust, adaptive defenses against even novel attack strategies.

The primary goal of this research is to design and implement an IDS that leverages machine learning algorithms to detect and classify black hole attacks in MANETs. To achieve this, we have created a custom dataset by selecting the most relevant features that effectively capture the characteristics of black hole attacks. Our approach not only aims to enhance the detection accuracy but also to provide a scalable solution that can be adapted to various MANET environments.

The rest of this paper is organized as follows: the second section reviews the related literature. In the third section, we discuss the application of machine learning algorithms for security. The fourth section presents our proposed IDS approach. The fifth section details our custom dataset. The sixth section discusses the results and provides an in-depth analysis. Finally, the paper concludes with a summary of our findings and suggestions for future work.

2. LITERATURE REVIEW

2.1. Related works

Various studies [8]–[12] have been conducted to enhance the security of MANETs, focusing on different approaches and methodologies. This section discusses the related work in this domain by summarizing key findings from previous research efforts. The review highlights the progress made in securing MANETs.

In one study by Almomani *et al.* [13], a comparison of various machine learning classifiers was conducted to identify the most effective one for intrusion detection in MANETs. The study found that the random forest (RF) classifier outperformed the others, achieving an accuracy rate of 87%. Additionally, the RF classifier demonstrated high F-measure and precision scores, indicating its robustness in detecting intrusions.

Kocher and Kumar [14] utilized machine learning models trained on the UNSW-NB15 dataset [15] to classify network traffic. The evaluation included k-nearest neighbors (k-NN), stochastic gradient descent (SGD), decision tree (DT), random forest (RF), logistic regression (LR), and naïve Bayes (NB) classifiers. Among these, the RF classifier achieved the highest performance, with an accuracy rate of 95.43%.

In another study, Saranya *et al.* [16] conducted a comprehensive survey on intrusion detection using machine learning algorithms. They evaluated the performance of linear discriminant analysis (LDA), classification and regression tree (CART), and RF techniques on the KDD'99 Cup dataset [17]. The results showed that the RF algorithm outperformed the other techniques, achieving an impressive accuracy of 99.81%, compared to LDA's 98.1% and CART's 98%. This study highlights the effectiveness of RF in intrusion detection tasks and its potential superiority over other commonly used algorithms.

Laqtib *et al* [18] conducted a systematic comparison of three intrusion detection systems based on the Inception architecture inception convolutional neural network (CNN), bidirectional long short-term memory (BLSTM), and deep belief network (DBN). They used the NSL-KDD dataset [19], which includes data on both intrusions and normal network connections. Their study aimed to offer foundational guidance on choosing appropriate deep learning models for MANET environments.

According to Abrar *et al.* in [20], various machine learning classifiers support vector machine (SVM), k-NN, LR, NB, multi-layer perceptron (MLP), RF, extra-tree classifier (ETC), and DT were utilized to classify data as normal or intrusive. Their model performance was evaluated using four feature subsets extracted from the NSL-KDD dataset. Empirical results showed that RF, ETC, and DT achieved performance above 99% for all attack classes across different feature subsets.

Alangari [21] proposed an advanced hybridized optimization technique (AHGFFA) that uses unsupervised machine learning to enhance security in MANET-internet of things (IoT) sensor systems. This method incorporates secure certificate-based group formation (SCGF) and recommended action K-means (K-RF means) filtering to organize the network into groups and optimize routing based on trust calculations. The combination of genetic and Firefly algorithms in AHGFFA ensures safe and efficient routing.

The study in [22] introduces a model that leverages the particle bee colony swarm (PBCS) algorithm for efficient routing and integrates the hybrid AdaBoost-RF algorithm to reduce training time while enhancing accuracy. To prevent attacks, the model employs the ad hoc on-demand distance vector (AODV) protocol, which is a widely used approach in MANETs. The model's performance is thoroughly evaluated using various metrics, including accuracy, recall, precision, energy consumption, and network lifetime, demonstrating its effectiveness in both security and resource management.

These studies provide valuable insights into the application of machine learning techniques to enhance the security of MANETs. They demonstrate the effectiveness of various classifiers and detection methods, showcasing the potential of these approaches in addressing security challenges. Building upon this existing research, our study makes a significant contribution by developing an intrusion detection system with a custom dataset specifically designed to detect black hole attacks in MANETs. This tailored dataset enables a comprehensive evaluation of machine learning models, leading to improved accuracy and reliability in intrusion detection. By focusing on the unique characteristics of black hole attacks, our approach advances the current state of MANET security and offers a robust framework for future research and implementation.

2.2. Black hole attack in MANET

In a black hole attack, a malicious node manipulates the routing protocol to attract all network traffic towards itself, ultimately dropping the packets. This action creates a “black hole” in the network, where data is absorbed and lost, disrupting communication [4], [23]. The attack exploits the trust-based nature of routing protocols, particularly in MANETs.

Specifically, the malicious node sends a false route reply (RREP) packet to the source node, falsely claiming that it has the shortest path to the destination. This deceitful action takes advantage of protocols like AODV [24], where nodes depend on received route replies to establish paths for data transmission. The source node, unaware of the malicious intent, then forwards all its data packets to the attacker. The attacker, instead of forwarding these packets, drops them or keeps them, effectively isolating the source node from the rest of the network and causing a breakdown in communication. The mechanism of the black hole attack is illustrated in Figure 1.

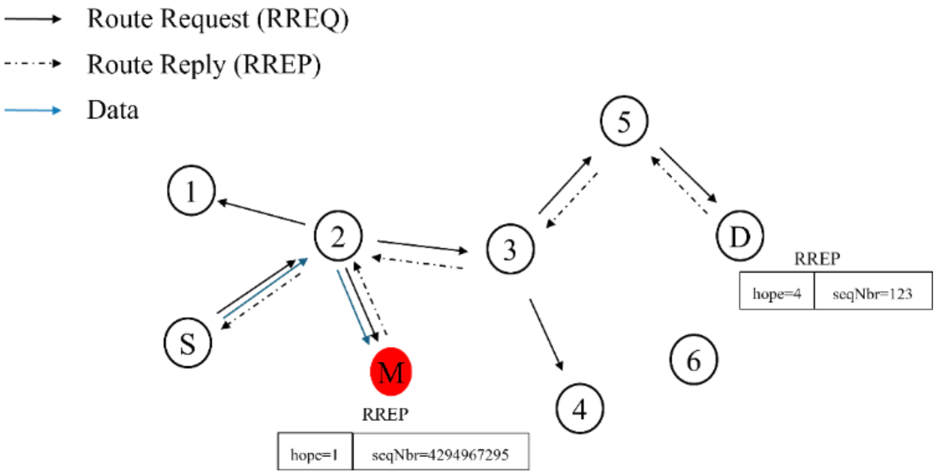


Figure 1. The black hole attack

3. MACHINE LEARNING ALGORITHMS FOR SECURITY

LR [25] is a statistical method used for predictive modeling. It is a generalized linear model that is used to model the relationship between a binary outcome variable and one or more independent variables. The outcome variable is a binary variable (also known as a dependent variable) that can take on one of two possible values, such as “yes” or “no”, “true” or “false”, or “1” or “0” as shown in Figure 2.

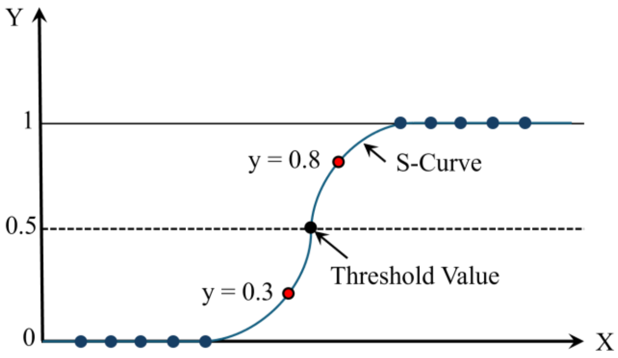


Figure 2. Logistic regression (LR)

DT classifier [26] is a supervised learning algorithm used in classification problems. It creates a tree-like model of decisions and their possible outcomes, including predictions of the output value. The tree is constructed by recursively splitting the data into subsets based on the values of the input features, with each split leading to a specific decision or prediction. The final predictions are made by traversing the tree from the root to a leaf node. The goal is to create a model that correctly classifies the majority of the training examples and generalizes well to new, unseen examples. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome as shown in Figure 3.

RF [27] is an ensemble learning method used for classification, regression, and other tasks. It constructs multiple DT during training and outputs the class that is the mode of the classes (for classification) or the mean prediction (for regression) of the individual trees. As a result of this combination, the forest usually achieves a better predictive accuracy than any individual tree. The key difference between a DT and a RF is that in a RF, a random subset of the features is chosen for each split, in addition to a random sample of the data. This introduces randomness into the model, which helps to reduce overfitting and improve the overall accuracy of the model. RF is based on the concept of ensemble learning, which involves combining multiple classifiers to solve a complex problem and improve the performance of the model as shown in Figure 4.

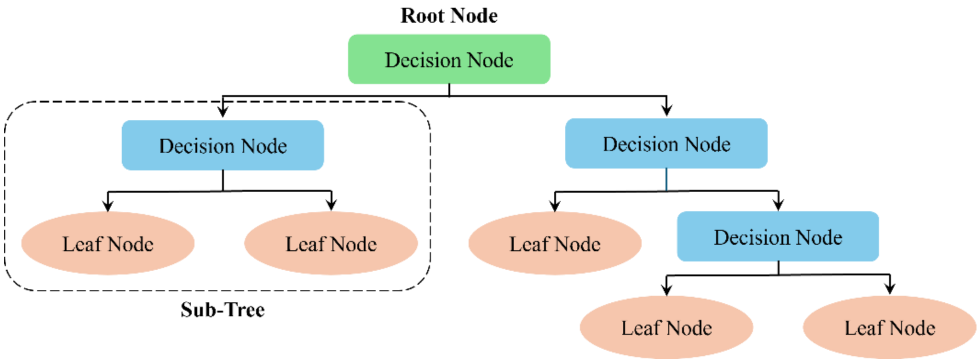


Figure 3. Decision tree

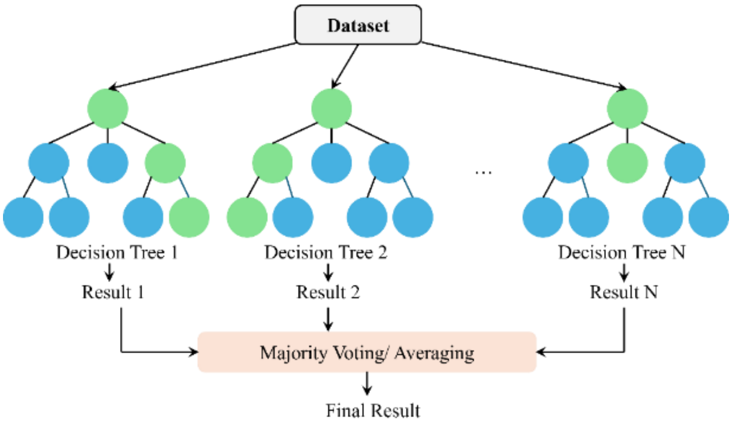


Figure 4. Random forest (RF)

k-NN [28] is an instance-based, or memory-based, supervised learning algorithm. The basic idea is to classify a new, unseen observation by determining the majority class of its “k” closest “neighbors” in the feature space. The neighbors are defined by a distance metric, such as Euclidean distance, which measures the similarity between two observations. The value of “k” is chosen by the user, and a common choice is to use k=5 or k=10. The algorithm is simple to implement but can be computationally expensive when working with large datasets. Additionally, it is sensitive to the choice of distance metric and the value of “k” as shown in Figure 5.

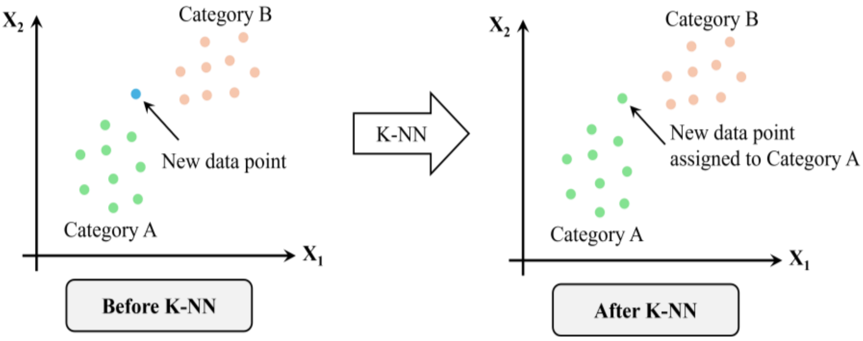


Figure 5. K-nearest neighbors (k-NN)

4. PROPOSED APPROACH

Our proposed system presents a novel approach for an intrusion detection system using LR, DT, RF, and k-NN algorithms, implemented in Python (version 3.12.3) [29]. The system architecture consists of six distinct steps:

Step 1. Establishing mobility within the NS-2 software [30] to generate realistic scenarios. This involves configuring the movement patterns of nodes within the network to mimic real-world conditions, such as varying speeds, pauses, and trajectories. By accurately simulating node mobility, we can better understand how the network behaves under different circumstances and how it is affected by potential intrusions.

Step 2. Generating both normal and malicious nodes in the MANET using the NS-2 software, resulting in trace files and network animator (NAM) files. The trace file captures the required dataset, while the NAM file describes node communication.

Step 3. Extracting observations from the NS-2 trace file to serve as inputs for the system using an AWK script. This extraction process ensures that all relevant metrics are captured, providing a dataset for subsequent analysis and modeling within the intrusion detection system.

Step 4. Preprocessing the data involves converting categorical values to numerical ones, cleaning the data by removing rows with missing or infinite values, and eliminating duplicate rows. We balance the dataset using random undersampling. Outliers are then removed from all the features, and the features are standardized. Finally, our dataset is divided into training and test sets.

Step 5. Training the IDS system using LR, DT, RF, and K-NN algorithms with the training set of our dataset.

Step 6. Testing the system's performance by evaluating detection accuracy using different confusion matrices and assessing precision, recall, and F1-score.

To provide a clearer understanding of our proposed method, the following flowchart in Figure 6 shows the step-by-step process of our intrusion detection system, from scenario generation to model training and evaluation.

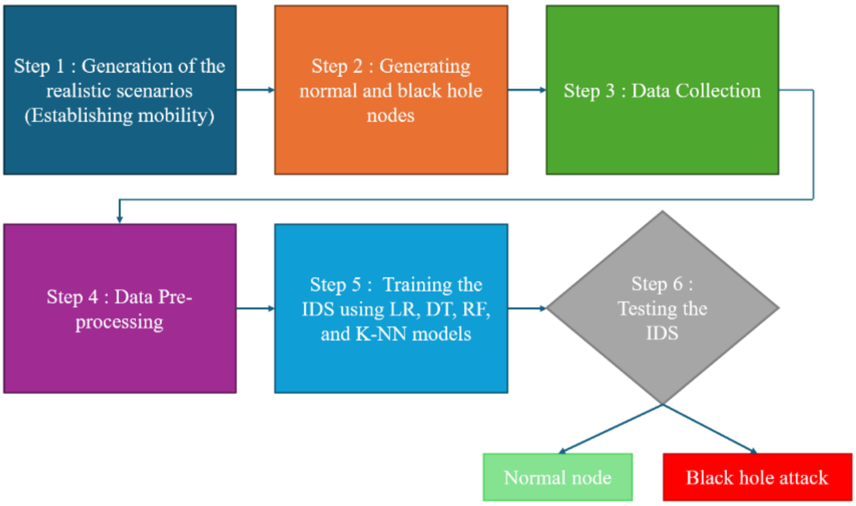

Step 1 : Generation of the realistic scenarios (Establishing mobility)
Step 2 : Generating normal and black hole nodes
Step 3 : Data Collection
Step 4 : Data Pre-processing
Step 5 : Training the IDS using LR, DT, RF, and K-NN models
Step 6 : Testing the IDS
Normal node
Black hole attack

Figure 6. Flowchart of the proposed intrusion detection system

5. OUR DATASET

5.1. Data collection and preparation

In the data collection and preparation step, we discussed the tools to be used and the creation of scenarios. This involved using specific tools and techniques to collect data and prepare it for further analysis and modeling. Scenarios were created to simulate real-world conditions and capture different behaviors and interactions in a MANET.

Throughout the data collection process, a range of tools were employed to gather pertinent information, including packet transmission data, network topology, and network performance metrics. By utilizing these tools, the necessary data points for analysis were effectively captured. In addition, scenarios were designed and simulated to mimic real-world situations, including different network conditions, node behaviors, and black hole attacks. This involved setting up the network environment, configuring node behaviors, and generating data through simulations.

The simulations were conducted using NS-2 (v-2.35) network simulator to evaluate the performance of our proposed solution against black hole nodes. In a 520x520 m area, 25 nodes were randomly distributed. They executed the AODV routing protocol both without attack and under the black hole attack scenario. Malicious nodes were also randomly distributed. Ten pairs were selected randomly for data communication, each transmitting at 512 bytes per second. All nodes moved according to a random waypoint model with speeds ranging randomly from 0 to 30 m/s. Additionally, nodes had a pause time of 10 seconds. Table 1 provides a summary of the simulation parameters.

Once the data was collected, it underwent a series of preparation steps to ensure its quality and suitability for analysis. These steps included cleaning the data by removing any irrelevant or erroneous entries, managing missing values, and converting the data into an appropriate format for analysis. This thorough preparation was essential to ensure the accuracy and reliability of the subsequent analysis.

Overall, the data collection and preparation process involved the use of appropriate tools to create realistic scenarios and accurately capture the data. By carefully processing the collected data, we ensured that it was ready for comprehensive analysis. This meticulous approach laid a strong foundation for the study's findings.

Table 1. Simulation parameters

Parameter	Value
Coverage area	520×520 m
Number of nodes	25
Simulation time	200 s
Transmission range	50 m
Mobility model	Random way point
Data rate	0.25
Packet size	512 Bytes
Routing protocol	AODV
Mobility speed	0-30 m/s
No of black hole nodes	5
Connections	10
Traffic type	UDP–CBR
Pause time	10 s

5.2. Feature selection and feature extraction

During the feature selection and extraction step, we analyzed the black hole attack in detail to identify key indicators for effective detection. Using an awk script, we parsed and processed the data to extract relevant features. This process transformed raw data into a refined set of features that capture essential information about the attack. These features are crucial for building an effective detection model, as they provide valuable insights for accurate classification and detection of the attack.

We used the following features to detect the black hole attack in MANET: packets sent, packets received, packets forwarded, packets dropped, sent route request (RREQ), received RREQ, sent route reply (RREP), and energy left. Each of these features was carefully chosen based on its relevance to identifying the black hole attack. Below is a detailed explanation of each feature's importance.

a. Packets sent: monitoring the packets sent by a node helps us gauge its activity level. In the context of a black hole attack, the malicious node may fail to forward packets it receives, leading to a noticeable decrease in packet transmission compared to the overall network traffic. This helps in identifying nodes that are behaving abnormally, as legitimate nodes should have a consistent rate of sent packets.

b. Packets received: analyzing the number of packets received allows us to identify nodes that exhibit unusual behavior. A selective black hole node drops or ignores packets, leading to a reduced number of received packets. This feature is crucial for detecting discrepancies in packet reception, which can signal malicious activity.

c. Packets forwarded: this feature enables us to observe the packet forwarding behavior of nodes in the network. A black hole node typically avoids forwarding packets to legitimate destinations, resulting in a lower number of packets forwarded compared to other nodes. This feature is directly related to the core behavior of a black hole attack, where the malicious node disrupts normal data flow.
d. Packet dropped: tracking the number of dropped packets is essential for identifying black hole attacks. Malicious nodes tend to intentionally drop packets, disrupting the normal flow of communications. An increased number of dropped packets can indicate the presence of a black hole node, as legitimate nodes should maintain a low drop rate.
e. Forward packet ratio: this feature measures the proportion of packets that a node forwards compared to the packets it receives. It indicates the efficiency and reliability of a node in forwarding data within the network. A high forward packet ratio suggests effective packet forwarding, while a low ratio may indicate issues such as packet loss or malicious activity like a black hole attack.
f. Sent RREQ: analyzing the number of RREQ packets sent by a node helps us determine if a node is actively involved in route discovery. This feature helps identify nodes that may be fabricating routes, which is a tactic used in black hole attacks.
g. Received RREQ: monitoring the reception of RREQ packets allows us to evaluate the participation of nodes in the route discovery process.
h. Sent RREP: tracking the number of RREP packets generated by nodes helps identify nodes that may be fabricating or modifying RREP packets to mislead the network. A black hole node can fabricate or modify RREP packets to attract traffic towards itself, causing abnormal patterns in sent RREP packets.
i. Energy left: monitoring the energy level of nodes is crucial for detecting a black hole attack. Malicious nodes may consume energy more rapidly due to their malicious activities, such as generating unnecessary traffic or performing additional operations. A rapid decline in energy can indicate malicious behavior.

6. RESULT AND DISCUSSION

6.1. Dataset statistics overview

Figure 7 depicting the statistics of our dataset, including the distribution of all the studied features. This visualization provides an overview of key measures such as mean, standard deviation, median, minimum, and maximum values, as well as quartiles. This graphical representation helps in understanding the distribution and trends of the data in our dataset, thereby facilitating the analysis and interpretation of the results.

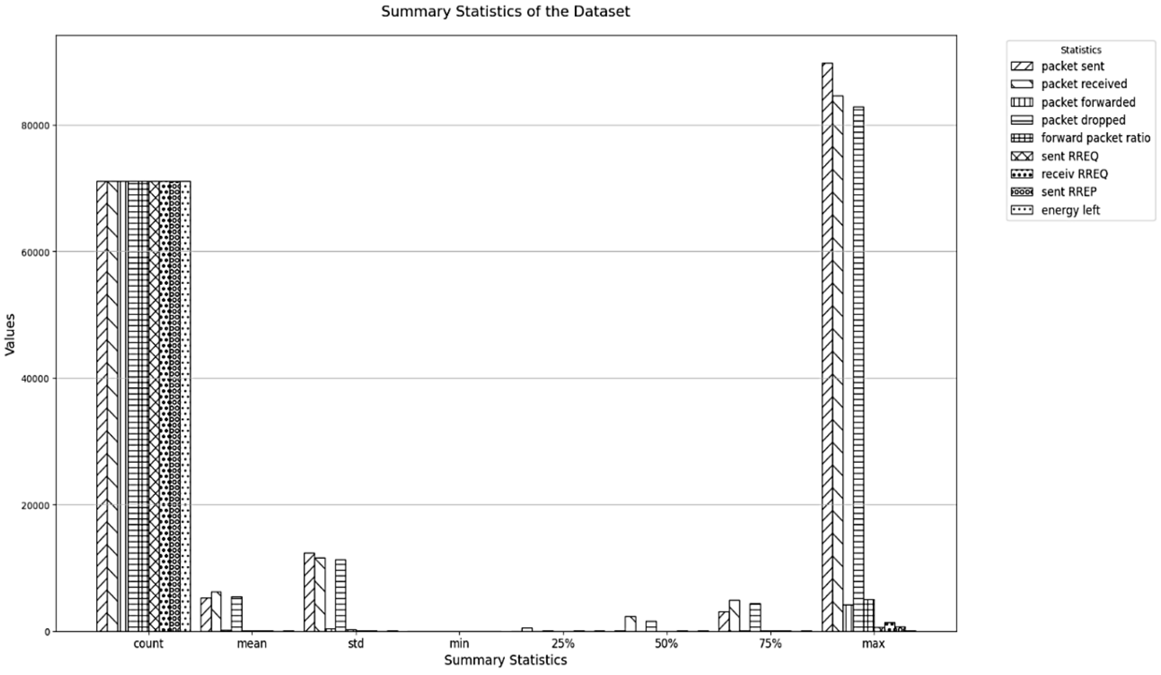
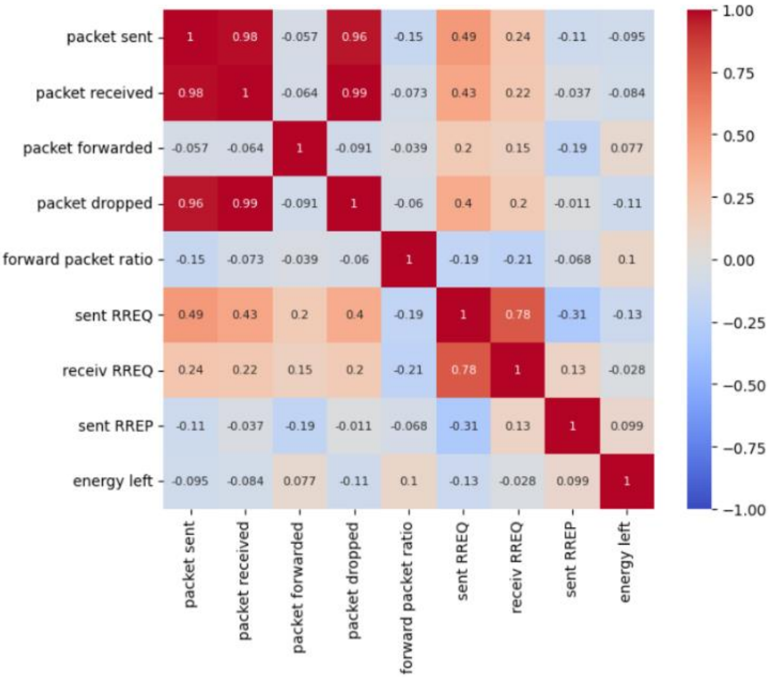


Figure 7. Summary statistics of our dataset

6.2. Correlation matrix

The correlation matrix displays the correlation coefficients between different variables associated with packet transmission in a network. The values range from -1 to 1. A positive correlation indicates that the two variables tend to increase or decrease together, while a negative correlation means that as one variable increases, the other decreases, and vice versa. A value of 0 implies no correlation. The correlation matrix of our dataset is shown in Figure 8. Here are some important findings:

a. The variables "packet sent," "packet received," and "packet dropped" are highly correlated with each other (all values are above 0.96). This means that when more packets are sent, there is a higher likelihood of receiving and dropping them.
b. The variable "packet forwarded" is not correlated with "packet sent" and "packet received." This implies that not all sent or received packets are being forwarded.
c. The "forward packet ratio" has a slight negative correlation with "packet sent," "packet received," and "packet dropped." This suggests that as the number of sent, received, or dropped packets increases, the forward packet ratio decreases.
d. The variables "sent RREQ" (route request) and "receive RREQ" (received route request) are positively correlated, as expected.
e. The "energy left" variable shows a negative correlation with packet-related variables (packet sent, packet received, packet dropped), suggesting that more transmission activity results in less remaining energy.



	packet sent	packet received	packet forwarded	packet dropped	forward packet ratio	sent RREQ	receiv RREQ	sent RREP	energy left
packet sent	1	0.98	-0.057	0.96	-0.15	0.49	0.24	-0.11	-0.095
packet received	0.98	1	-0.064	0.99	-0.073	0.43	0.22	-0.037	-0.084
packet forwarded	-0.057	-0.064	1	-0.091	-0.039	0.2	0.15	-0.19	0.077
packet dropped	0.96	0.99	-0.091	1	-0.06	0.4	0.2	-0.011	-0.11
forward packet ratio	-0.15	-0.073	-0.039	-0.06	1	-0.19	-0.21	-0.068	0.1
sent RREQ	0.49	0.43	0.2	0.4	-0.19	1	0.78	-0.31	-0.13
receiv RREQ	0.24	0.22	0.15	0.2	-0.21	0.78	1	0.13	-0.028
sent RREP	-0.11	-0.037	-0.19	-0.011	-0.068	-0.31	0.13	1	0.099
energy left	-0.095	-0.084	0.077	-0.11	0.1	-0.13	-0.028	0.099	1

Figure 8. Correlation matrix

6.3. Evaluation of machine learning models

The confusion matrix for the four models-logistic regression, DT, RF, and k-NN-are presented in Tables 2, 3, 4, and 5, respectively. Table 6 provides the evaluation metrics for the four machine learning models applied to classify black hole attacks using the customized MANET dataset. The Figure 9 showcases the performance of four machine-learning models for our binary-class classification problem. The results display the best scores and test scores for each model. For LR, the highest score achieved was 0.964, with a test score of 0.951, demonstrating good performance on unseen data. The DT model attained high scores, with a best score of 0.971 and a test score of 0.9565, demonstrating excellent performance on the test data. The RF model performed well, with a best score of 0.986 and a test score of 0.977. This indicates that the model generalizes effectively to previously unseen data. The k-NN model achieved a best score of 0.955 and a test score of 0.949. Although the test score is slightly lower, it still demonstrates satisfactory performance. Overall, these results highlight the effectiveness of machine learning models in data classification. The DT and RF models show particularly strong performance, while logistic regression and k-NN also provide respectable results.

Table 2. Confusion matrix for LR

	Predicted normal	Predicted attack
Actual normal	33,909	1,641
Actual attack	1,394	32,674

Table 3. Confusion matrix for DT

	Predicted normal	Predicted attack
Actual normal	34,280	1,270
Actual attack	1,079	33,360

Table 4. Confusion matrix for RF

	Predicted normal	Predicted attack
Actual normal	34,695	855
Actual attack	709	34,140

Table 5. Confusion matrix for k-NN

	Predicted normal	Predicted attack
Actual normal	33,826	1,724
Actual attack	1,516	32,501

Table 6. Evaluation metrics for machine learning models

Model	Accuracy	Precision	Recall	F1-score
LR	0.960	0.955	0.959	0.957
DT	0.970	0.965	0.968	0.968
RF	0.980	0.975	0.979	0.977
k-NN	0.955	0.950	0.955	0.953

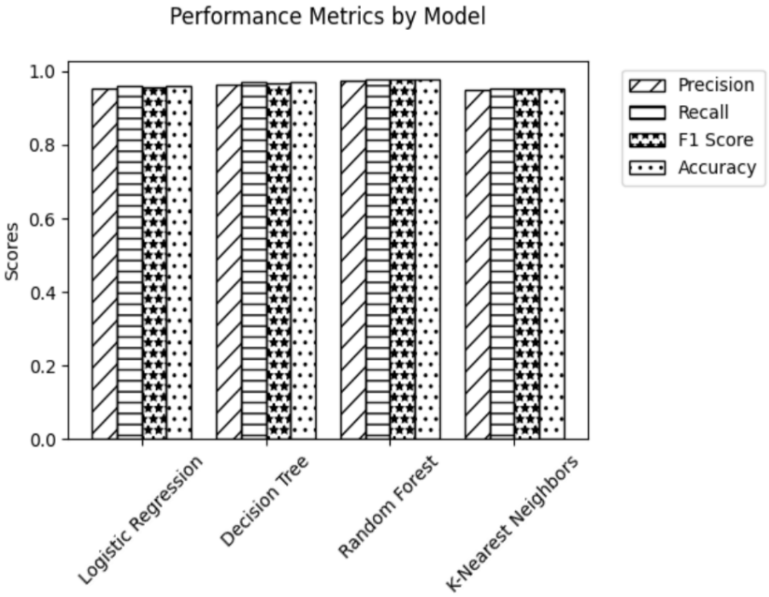


Figure 9. Performance metrics by model

6.4. Discussion of the results

The performance evaluation of the machine learning models in our study demonstrates that RF, logistic regression, k-nearest neighbors, and DT models can classify black hole attacks in MANETs with a reasonable degree of accuracy. In particular, the DT and RF models achieved the highest scores, indicating their strong ability to generalize to unseen data. These results underscore the potential of these models to effectively detect black hole attacks, aligning with our primary objective to enhance MANET security.

Our study's main contribution lies in the use of a uniquely collected dataset tailored to the characteristics of black hole attacks in MANETs. Unlike previous studies that often relied on publicly available datasets, our dataset provides a more accurate reflection of the conditions and challenges specific to MANET environments. This has allowed us to obtain results that are more applicable to real-world scenarios. The strength of our approach is not only in the selection of effective machine learning models, such as DT and RF, but also in the relevance and specificity of our data, which enhances the models' performance and reliability. However, a limitation of our study is that the dataset may not encompass all possible variations of attack patterns. Future work could involve expanding the dataset with more diverse scenarios or combining it with other datasets to further improve the robustness of the models.

In summary, our study provides valuable insights into the application of machine learning models, particularly DT and RF, for detecting black hole attacks in MANETs. These findings contribute to the broader field of network security by demonstrating the feasibility of deploying machine learning-based intrusion detection systems in resource-constrained environments. Future research could expand on this work by exploring the integration of these models with other security mechanisms in MANETs, as well as investigating their performance in more diverse and dynamic network scenarios.

7. CONCLUSION

In conclusion, our proposed approach to designing an intrusion detection system for black hole attacks in MANETs has demonstrated its effectiveness through comprehensive data collection, feature selection, and model evaluation. The results obtained from training and evaluating various machine learning models, including RF, logistic regression, k-nearest neighbors, and DT, underscore the potential of these models in accurately classifying black hole attacks. Particularly, the DT and RF models excelled, showcasing their ability to generalize well to unseen data, which is crucial for real-world deployment.

The findings of our study have significant implications for the field of network security, particularly in the context of MANETs. By demonstrating the feasibility of applying machine learning algorithms to detect black hole attacks, we contribute to the ongoing efforts to enhance the security and resilience of these networks. This research lays the groundwork for developing more sophisticated intrusion detection systems that can adapt to the evolving nature of network threats.

Moving forward, our work opens several avenues for future research. One potential application is the integration of our intrusion detection system with other security mechanisms, such as encryption and anomaly detection, to create a more comprehensive defense strategy for MANETs. Additionally, expanding the dataset to include a broader range of attack scenarios and testing the models in more dynamic and large-scale environments could further validate and enhance the robustness of our approach. In summary, our findings provide a solid foundation for further research and development in securing MANETs, with the potential to significantly impact the field and contribute to the broader community's efforts in safeguarding critical communication infrastructures.

REFERENCES

[1] D. Ramphull, A. Mungur, S. Armoogum, and S. Pudaruth, “A review of mobile ad hoc network (MANET) protocols and their applications,” in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, May 2021, pp. 204–211, doi: 10.1109/ICICCS51141.2021.9432258.

[2] M. AlRubaiei, H. sh Jassim, B. T. Sharef, S. Safdar, Z. T. Sharef, and F. L. Malallah, “Current vulnerabilities, challenges and attacks on routing protocols for mobile ad hoc network: a review,” in *Swarm Intelligence for Resource Management in Internet of Things*, Elsevier, 2020, pp. 109–129, doi: 10.1016/B978-0-12-818287-1.00012-7.

[3] B. Banerjee and S. Neogy, “A brief overview of security attacks and protocols in MANET,” in *2021 IEEE 18th India Council International Conference (INDICON)*, IEEE, Dec. 2021, pp. 1–6, doi: 10.1109/INDICON52576.2021.9691554.

[4] H. Moudni, M. Er-Rouidi, H. Mouncif, and B. El Hadadi, “Performance analysis of AODV routing protocol in MANET under the influence of routing attacks,” in *2016 International Conference on Electrical and Information Technologies (ICEIT)*, IEEE, May 2016, pp. 536–542, doi: 10.1109/EITech.2016.7519658.

[5] F. Hamza and S. Maria Celestin Vigila, “Review of machine learning-based intrusion detection techniques for MANETs,” in *Computing and Network Sustainability: Proceedings of IRSCNS 2018*, 2019, pp. 367–374, doi: 10.1007/978-981-13-7150-9_39.

[6] M. S. Hussain and K. U. R. Khan, “A survey of IDS techniques in manets using machine-learning,” in *Proceedings of the Third International Conference on Computational Intelligence and Informatics: ICCII 2018*, 2020, pp. 743–751.

[7] K. S and J. R. Prathuri, “Classification of misbehaving nodes in MANETS using machine learning techniques,” in *2020 2nd PhD Colloquium on Ethically Driven Innovation and Technology for Society (PhD EDITS)*, IEEE, Nov. 2020, pp. 1–2, doi: 10.1109/PhDEDITS51180.2020.9315311.

[8] H. Moudni, M. Er-Rouidi, H. Mouncif, and B. El Hadadi, “Fuzzy logic based intrusion detection system against black hole attack in mobile ad hoc networks,” *International Journal of Communication Networks and Information Security*, vol. 10, no. 2, pp. 366–373, 2018, doi: 10.17762/ijcnis.v10i2.3267.

[9] J. Y. Hande and R. Sadiwala, “Data security-based routing in MANETs using key management mechanism,” *SN Computer Science*, vol. 5, no. 1, p. 155, Jan. 2024, doi: 10.1007/s42979-023-02409-5.

[10] H. Moudni, M. Er-Rouidi, H. Mouncif, and B. El Hadadi, “Modified AODV routing protocol to improve security and performance against black hole attack,” in *2016 International Conference on Information Technology for Organizations*

Development (IT4OD), IEEE, Mar. 2016, pp. 1–7, doi: 10.1109/IT4OD.2016.7479265.

[11] H. Moudni, M. Er-Rouidi, H. Faouzi, H. Mouncif, and B. El Hadadi, "Enhancing security in optimized link state routing protocol for mobile ad hoc networks," in *Ubiquitous Networking: Third International Symposium, UNet 2017, Casablanca, Morocco, May 9-12, 2017, Revised Selected Papers 3*, 2017, pp. 107–116.

[12] J. A. Rathod and M. Kotari, "Secure and efficient message transmission in MANET using hybrid cryptography and multipath routing technique," *Multimedia Tools and Applications*, Jun. 2024, doi: 10.1007/s11042-024-19542-9.

[13] O. Almomani, M. A. Almaiah, A. Alsaaidah, S. Smadi, A. H. Mohammad, and A. Althunibat, "Machine learning classifiers for network intrusion detection system: comparative study," in *2021 International Conference on Information Technology (ICIT)*, IEEE, Jul. 2021, pp. 440–445, doi: 10.1109/ICIT52682.2021.9491770.

[14] G. Kocher and G. Kumar, "Performance analysis of machine learning classifiers for intrusion detection using unsw-nb15 dataset," *Comput. Sci. Inf. Technol.(CS IT)*, vol. 10, no. 20, pp. 31–40, 2020.

[15] V. Kumar, A. K. Das, and D. Sinha, "Statistical analysis of the UNSW-NB15 dataset for intrusion detection," in *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*, 2020, pp. 279–294, doi: 10.1007/978-981-13-9042-5_24.

[16] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. K. A. A. Khan, "Performance analysis of machine learning algorithms in intrusion detection system: a review," *Procedia Computer Science*, vol. 171, pp. 1251–1260, 2020, doi: 10.1016/j.procs.2020.04.133.

[17] S. Kumar, Sunanda, and S. Arora, "A statistical analysis on KDD Cup'99 dataset for the network intrusion detection system," *Applied Soft Computing and Communication Networks: Proceedings of ACN 2019*, pp. 131–157, 2020, doi: 10.1007/978-981-15-3852-0_9.

[18] S. Laqtib, K. El Yassini, and M. L. Hasnaoui, "A technical review and comparative analysis of machine learning techniques for intrusion detection systems in MANET," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, pp. 2701-2709, Jun. 2020, doi: 10.11591/ijece.v10i3.pp2701-2709.

[19] R. Bala, "A review on KDD CUP99 and NSL-KDD datset," *International Journal of Advanced Research in Computer Science*, vol. 10, no. 2, pp. 64–67, Apr. 2019, doi: 10.26483/ijarcs.v10i2.6395.

[20] I. Abrar, Z. Ayub, F. Masoodi, and A. M. Bamhdi, "A Machine learning approach for intrusion detection system on NSL-KDD dataset," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, Sep. 2020, pp. 919–924, doi: 10.1109/ICOSEC49089.2020.9215232.

[21] S. Alangari, "An unsupervised machine learning algorithm for attack and anomaly detection in IoT sensors," *Wireless Personal Communications*, Feb. 2024, doi: 10.1007/s11277-023-10811-8.

[22] S. Sophie Maria Vincent and N. Duraipandian, "Detection and prevention of sinkhole attacks in MANETS based routing protocol using hybrid adaboost-random forest algorithm," *Expert Systems with Applications*, vol. 249, p. 123765, Sep. 2024, doi: 10.1016/j.eswa.2024.123765.

[23] M. M. Hamdi *et al.*, "A study review on gray and black hole in mobile ad hoc networks (MANETs)," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022, pp. 1–6, doi: 10.1109/HORA55278.2022.9800011.

[24] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," RFC 3561, 2003, doi: 10.17487/rfc3561.

[25] S. Menard, *Applied logistic regression analysis*. United Kingdom: Sage, 2002.

[26] S. Suthaharan and S. Suthaharan, "Decision tree learning," *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, pp. 237–269, 2016.

[27] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017, doi: 10.17849/insm-47-01-31-39.1.

[28] L. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009, doi: 10.4249/scholarpedia.1883.

[29] W. McKinney, *Python for data analysis*. Sevastopol: O'Reilly Media, Inc., 2022.

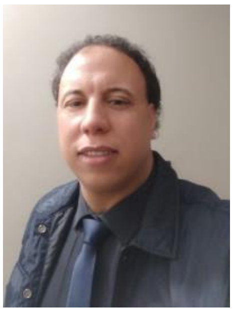
[30] T. Issariyakul and E. Hossain, *Introduction to network simulator NS2*. Boston: Springer US, 2009, doi: 10.1007/978-0-387-71760-9.

BIOGRAPHIES OF AUTHORS

Houda Moudni is currently working as an assistant professor at the National School of Business and Management, Sultan Moulay Slimane University, Béni Mellal, Morocco. She received the Ph.D. degree in computer sciences from the Faculty of Sciences and Technology of Béni Mellal in 2019. She developed a strong interest in computer networking. Her research work primarily focuses on securing routing protocols in mobile ad hoc networks (MANET), wireless sensor networks (WSN), and the internet of things (IoT). She can be contacted at email: h.moudni@usms.ma.

Mohamed Er-Rouidi is a dedicated researcher and a teacher-researcher at the multidisciplinary faculty of Cadi Ayyad University in Safi, Morocco. Holding a Ph.D. in computer science from the Faculty of Sciences and Techniques of Beni Mellal in 2019, he developed a strong interest in ad hoc networks. His research work primarily focuses on enhancing routing protocols and minimizing energy consumption in mobile ad hoc networks. His areas of expertise include mobile ad hoc networks as well as the internet of things (IoT), ever-evolving fields that require a deep understanding to address contemporary challenges. He can be contacted at email: mohamed.errouidi@uca.ac.ma.

Mansour Lmkaiti is a Ph.D. student in the Department of Computer Mathematics at the Polydisciplinary Faculty, Sultan Moulay Slimane University, Morocco. His domains of interest include high-performance computer systems and networks, machine learning algorithms, high performance in wireless sensor networks (WSNs), and cybersecurity in WSNs. He can be contacted at email: lamkaitimansour@gmail.com.



Hicham Mouncif Department of Computer Mathematics, University Sultan Moulay Slimane, Morocco. He is currently working as a professor at the Department of Mathematics and Informatics. His research interests include computer networking, communication engineering, and securing routing protocols in wireless sensor networks. Domains of interests: high-performance computer systems and networks: theory, machine learning algorithms; high performance in WSNs and cyber security. He can be contacted at email: h.mouncif@usms.ma.