

Black hole Attack Detection Using Machine Learning Algorithms in MANET – Performance Comparison

Ms. Katakam Tejaswini¹, Mrs. Yannam Adilakshmi²

¹Katakam Tejaswini Gudlavalleru Engineering College

²Mrs. Yannam Adilakshmi Associate Professor, Dept. of Computer Science and Engineering, Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India

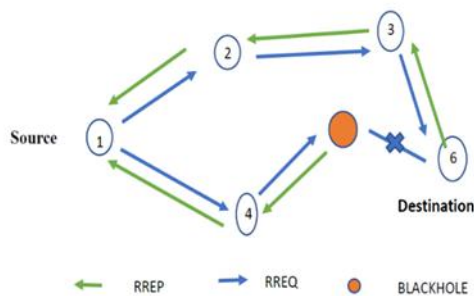
Abstract - MANET forms without any stand support devices. Nodes can be established in the network very easily without any prerequisite. Due to this flexibility many security threats can occur in the routing. So, in order to overcome from this, the performance of IDS should be improvised. In the present project a comparison study has been carried out on machine learning algorithms in terms of accuracy and detection rate for IDS improvisation.

Key Words: MANET, Support Vector Machine, Decision tree Classifier, Random Forest classifier, Logistic Regression

1. INTRODUCTION

IDS is to detect the attack before the attacker introduces any harm to the network, it take care of three major functions like monitoring, detecting and generating alarms. The black hole attack is considered as the one of the most affected kind on MANETs. In order to protect the network from blackhole attack, the use of an Intrusion Detection System (IDS) has a major importance in the MANET protection. With the help of the machine learning algorithms here it improve the anomaly-based IDS in MANET's here the following algorithms are support vector machine, decision tree classifier, random forest classifier, logistic regression algorithm by considering above four algorithms which one is best for detecting anomaly-based IDS in MANETs.

1.1 BLACK HOLE ATTACK USING AODV ROUTING PROTOCOL



Fig(1):Black hole Attack

Black hole attack is one of the major attacks in MANETs. It consists of Source node, Destination node and Neighbouring nodes. The source node sends a RREQ (Route Request Packet) to its neighbouring nodes to search for the route destination. However, black hole node sends a fake route reply to the source node which will degrade the performance of the network. In order to prevent this, the performance of the IDS should be increased with machine learning algorithms.

2. Related Work:

Sujithra L et. al In this paper it improves the conversation of energy in heterogenous network and also reducing the active time of IDS running in the nodes so in order to achieve this proabilistic approach is implemented, here optimal proabilistic of node to be set. By considering this decreasing the active time of IDS in each node and also conserve the energy of the node that increases the network lifetime significantly.

Sankaranarayanan.S et. al Proposed the RSA algorithm in intrusion detection system in MANET it successfully identifies the malicious node Results shows that secure IDS method gives better packet deliver ratio in presence of intruders.

Indira N Proposed Anomaly based intrusion detection technique using the SOM classification method provides higher detection rate than other anomaly detection method. As anomaly-based intrusion detection techniques are based on statistical data they can result in false positive identification of normal pattern as an attack. This false identification of benign behavior as abnormal can result in isolation of non-malicious node as malicious, thus may result in partitioning of the network.

Hangmei Deng proposed two detection models that is distributed hierarchical and completely distributed. IDS is implement- ed in both models and focus on the network layer for detect the attack in the network that hierarchical distribution gets the good detection rate than the complete distributed approach

Houda Moudni proposed-Adaptive neuro fuzzy inference system (ANFIS) algorithm and particle swarm optimization

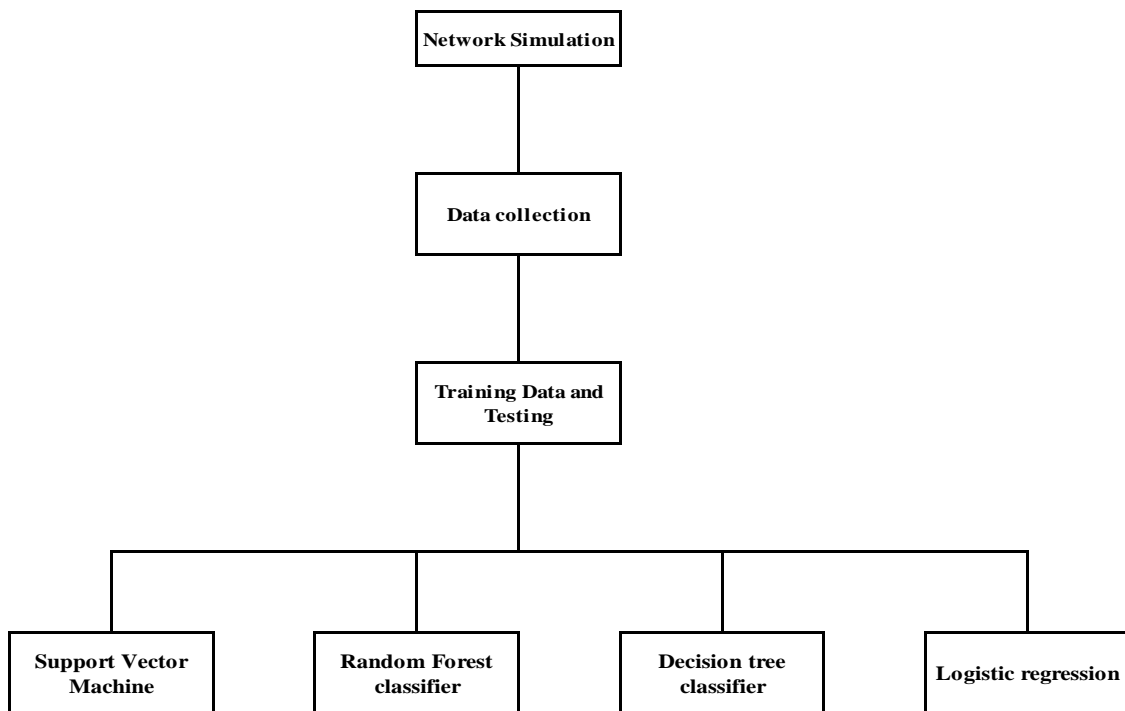
(PSO) are used for detection and prevention of black hole attack in MANETs. The performance evaluation (ANFIS+PSO) is evaluated by Detection rate and False rate. As a result; less no. of node connections leads to high detection rate and low false rate.

Zalte S.S. et. al proposed- black hole and gray hole attack detection in AODV protocol by IDS. By comparing the normal AODV, attacker AODV and new AODV. As a result, new AODV improves the throughput and packet delivery ratio.

3. PROPOSED METHODOLOGY

Nodes in the MANET share the wireless medium and the topology of the network changes erratically and dynamically. Research in a MANET gets tremendous attention because of its eminent characteristics like instant infrastructure, cheap

and easy deployment in hostile terrain where geographical conditions are not suitable for an earthquake, battlefield. MANET can build anytime and anywhere. Since the nodes are roaming, the network topology varies rapidly. The remarkable advantages of MANETs such as multi hop, infrastructure less transmission etc., makes it as a best medium to networks. Though MANETs have surplus things, they have some security issues that will cause severe damages and loss in network. Random linking of mobile nodes leads to add malicious nodes in the network accidentally. To suspect and detect the malicious activity in the network, Intrusion Detection System (IDS) are implemented to analyze the behaviour of the neighbourhood nodes. To improving the anomaly based intrusion detection system in MANETs comparing the four algorithms i.e, support vector machine, decision tree classifier, Random forest classifier and logistic regression.



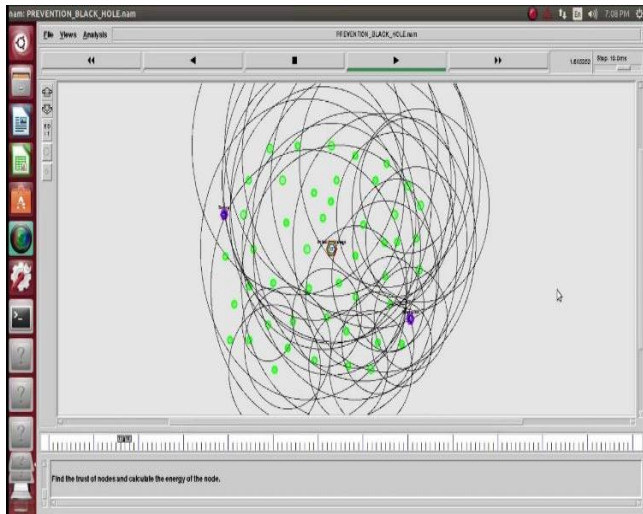
Fig(2):Flow chart for Proposed Methodology

A 3 step method is followed for the analysis

1. Simulation of Network
2. Data collection
3. Model Training and Data Testing

1. Simulation of Network

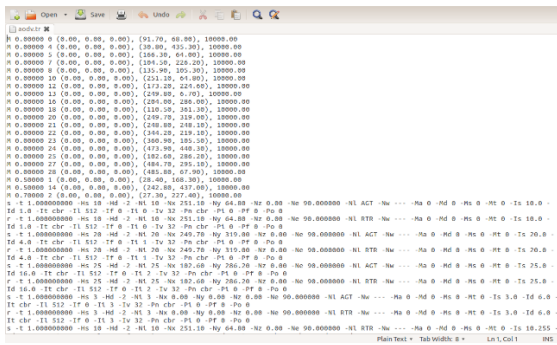
In the present project, Mobile ad-hoc network(MANETs) is simulated in NS2 with 25 number of nodes as shown in fig(8).



Fig(3) Simulation of Black Hole attack in ns2 tool

2. DATA COLLECTION:

After simulation, a trace file is generated from NS2 which will be an input for .CSV file. The output of trace file and input of .csv file are shown in Fig(4) and Fig(5) respectively. Generally, trace file has more number of attributes however, if the number of received packets are more than the number of dropped packets such kind of attributes have been selected as an input for .csv file.



Fig(4). Trace file generated from black hole attack simulation

Node	Number of generated packets	Number of sent packets	Number of received packets	Number of dropped packets	Drop Count	Difference	Drop Rate
0	4	2	0	2	0	2	0.500000
1	10	10	0	0	0	0	0.000000
2	10	0	0	10	10	0	1.000000
3	10	10	0	0	0	0	0.000000
4	200	200	0	0	0	0	0.000000
5	5	5	0	0	0	0	0.000000
6	10	10	0	0	0	0	0.000000
7	400	400	0	0	0	0	0.000000
8	15	15	2	13	0	13	0.866667
9	100	100	200	0	0	0	0.000000
10	5	5	5	0	0	0	0.000000
11	200	200	0	0	0	0	0.000000
12	200	200	0	0	0	0	0.000000
13	200	200	0	0	0	0	0.000000
14	200	200	0	0	0	0	0.000000
15	200	200	0	0	0	0	0.000000
16	200	200	0	0	0	0	0.000000
17	200	200	0	0	0	0	0.000000
18	200	200	0	0	0	0	0.000000
19	200	200	0	0	0	0	0.000000
20	200	200	0	0	0	0	0.000000
21	200	200	0	0	0	0	0.000000
22	200	200	0	0	0	0	0.000000
23	200	200	0	0	0	0	0.000000
24	200	200	0	0	0	0	0.000000
25	200	200	0	0	0	0	0.000000
26	200	200	0	0	0	0	0.000000
27	200	200	0	0	0	0	0.000000
28	200	200	0	0	0	0	0.000000
29	200	200	0	0	0	0	0.000000
30	200	200	0	0	0	0	0.000000
31	200	200	0	0	0	0	0.000000
32	200	200	0	0	0	0	0.000000
33	200	200	0	0	0	0	0.000000
34	200	200	0	0	0	0	0.000000
35	200	200	0	0	0	0	0.000000
36	200	200	0	0	0	0	0.000000
37	200	200	0	0	0	0	0.000000
38	200	200	0	0	0	0	0.000000
39	200	200	0	0	0	0	0.000000
40	200	200	0	0	0	0	0.000000
41	200	200	0	0	0	0	0.000000
42	200	200	0	0	0	0	0.000000
43	200	200	0	0	0	0	0.000000
44	200	200	0	0	0	0	0.000000
45	200	200	0	0	0	0	0.000000
46	200	200	0	0	0	0	0.000000
47	200	200	0	0	0	0	0.000000
48	200	200	0	0	0	0	0.000000
49	200	200	0	0	0	0	0.000000

Fig(5).Csv dataset generated from tracefile

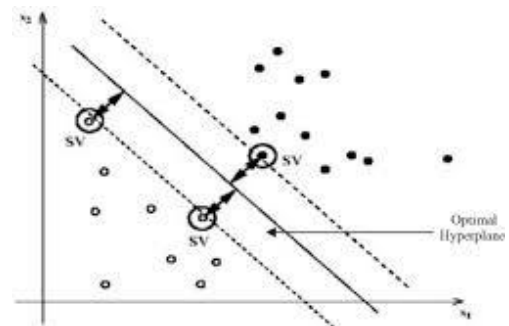
3. MODEL TRAINING AND DATA TESTING

In the present project four algorithms were used to train and test the data

- a. Support vector machine (SVM)
- b. Random forest classifier
- c. Decision tree classifier
- d. Logistic regression

a. Support vector machine

The primary aim of support vector machine(SVM) is to separate the normal and abnormal nodes by choosing the best estimated hyperplane and it is selected such away that the distance from the hyperplane to the nearest node on each side is maximized.



Fig(6).Support Vector Machine Classification

In the present Project, the data set obtained from NS2 is fed into SVM algorithm. The black hole attack is detected in terms of accuracy and confusion matrix. The output is shown in Fig (7).

```

File Edit View Insert Cell Kernel Help Not Traced [Python 3]
In [12]: from sklearn.svm import SVC
Out[12]: SVC(kernel='linear')

In [13]: svc.fit(x_train,y_train)
Out[13]: SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
decision_function_shape='raw', degree=3, gamma='auto_deprecated',
kernel='linear', max_iter=1, probability=False, random_state=None,
shrinking=True, tol=0.001, verbose=False)

In [14]: y_predict=svc.predict(y_test)

In [15]: y_predict
Out[15]: array([0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0])

In [16]: from sklearn.metrics import accuracy_score
Out[16]: accuracy_score(y_test,y_predict)
Out[16]: 0.8252941176470588

In [17]: cm=confusion_matrix(y_test,y_predict)

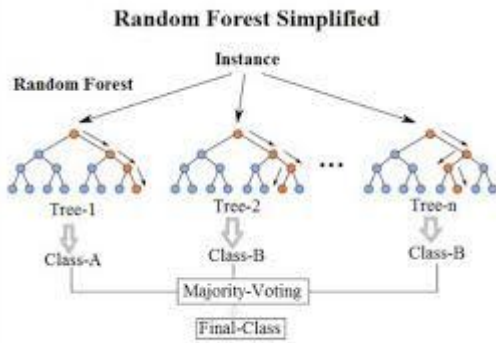
In [18]: cm
Out[18]: array([[13,  2],
               [ 1, 14]], dtype=int64)
    
```

Fig(7). Accuracy Score and Confusion Matrix of SVM Classifier

From SVM Algorithm it is observed that an accuracy of 82.53 and the confusion matrix showing less false positive rate

b. Random forest classifier

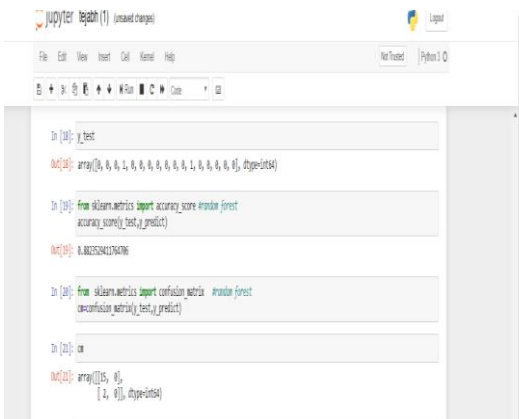
Random forest is an ensemble learning techniques used for predictive modelling and machine learning techniques. It classifies, create a set of decision tree and randomly selected subset of training set. It aggregates the votes from different decision tree to decide the final class of the test object.



Fig(8).Random Forest Classification

In the present Project, the data set obtained from NS2 is fed into Random forest classifier algorithm the black hole attack is detected in terms of accuracy and confusion matrix.

The output of Random forest classifier is shown in Fig(9).



Fig(9).Accuracy score and Confusion Matrix of Random Forest Classifier

From Random forest classifier algorithm, it is observed that an accuracy of 88.23and the confusion matrix showing less false positive rate

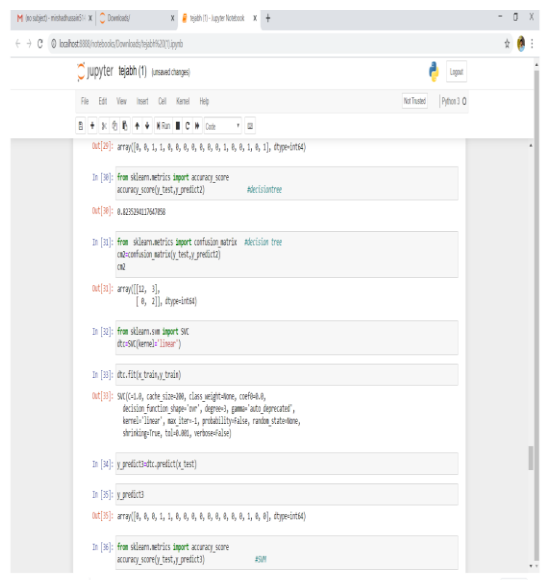
c.Decision Tree Classifier

Decision tree is a white box type of machine learning algorithms. It share internal decision making, logical which is not available in the black box type of algorithms . while building the decision tree the ID3 algorithm select the best feature at each step and it uses information gain to find best feature. Information gain is estimated using below formulae.

$$IG(D_p, f) = I(D_p) - \frac{N_{left}}{N} I(D_{left}) - \frac{N_{right}}{N} I(D_{right})$$

Where f - Feature to perform the split
 D p – Dataset of the parent node
 D left – Dataset of the left child node
 D right – Dataset of the right child node
 I – impurity criterion(Entropy)
 N – total number of samples
 N left – Number of samples at left child node
 N right – Number of samples at right child node
 In the present project, the data set obtained from NS2 is fed into Decision tree classifier algorithm the black hole attack is detected in terms of accuracy and confusion matrix

The output of Decision tree classifier is shown in Fig(10).



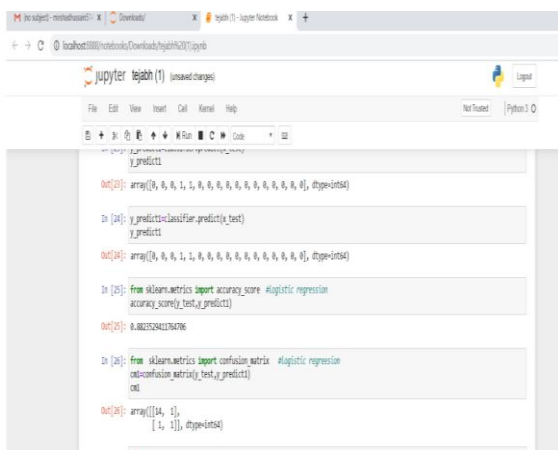
Fig(10). Accuracy score and Confusion Matrix of Decision Tree Classifier

From Decision Tree classifier algorithm, it is observed that an accuracy of 82.35 and the confusion matrix showing more false positive rate.

d.Logistic Regression

Logistic regression is one of the techniques for data classification. Categorical dependent variables can be solved by Logistic regression using some Mathematical models. In the present project, the data set obtained from NS2 is fed into Logistic Regression classifier algorithm the black hole attack is detected in terms of accuracy and confusion matrix.

The output of Logistic Regression is shown in Fig(11).



```

In [23]: y_predict

Out[23]: array([0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])

In [24]: y_predict_classifier.predict(x_test)
y_predict

Out[24]: array([0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])

In [25]: from sklearn.metrics import accuracy_score #logistic regression
accuracy_score(y_test,y_predict)

Out[25]: 0.882324179476

In [26]: from sklearn.metrics import confusion_matrix #logistic regression
cm=confusion_matrix(y_test,y_predict)
cm

Out[26]: array([[14, 1],
               [ 1, 1]], dtype=int64)
    
```

Fig(11). Accuracy score and Confusion Matrix of Logistic Regression Classifier

From Logistic Regression classifier algorithm, it is observed that an accuracy of 88.23 and the confusion matrix showing less false positive rate

By observing the above algorithms SVM and decision tree classifier obtained the same accuracy i.e, 82.35 , but SVM shows more detection rate by compare with randomforest and logistic regression getting same accuracy by comparing four algorithms random forest getting more detection rate.

By observing the above algorithms SVM and decision tree classifier obtained the same accuracy i.e, 82.35 , but SVM shows more detection rate by compare with randomforest and logistic regression getting same accuracy by comparing four algorithms random forest getting more detection rate.

S.No	Type of algorithm	Accuracy (%)	Detection rate
1	Support Vector Machine	82.35	Less
2	Random forest classifier	88.23	More
3	Decision tree classifier	82.35	Less
4	Logistic regression	88.23	Less

4 CONCLUSION

From above results it is concluded that random forest classifier gives the higher accuracy and better detection rate compared to the SVM algorithm, Decision tree classifier and Logistic regression. Therefore, the performance of IDS can be improvised by random forest algorithm

REFERENCES

1. International Research Journal of R & D IN ENGINEERING SCIENCE AND MANAGEMENT, Study of MANET-Characteristics, Challenges, Application, Routing Protocol and Security Attacks Sandeep Kumar, Col. (Dr) Suresh Kumar
2. International Conference on Advanced Computing and Communication Systems (ICACCS -2013), An Enhanced Intrusion Detection System for Routing Attacks in MANET2013 ,1 K Rama Abirami M G Sumithra, J Rajasekaran.
3. Neha Rai et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 4952-4957, Genetic Algorithm Based Intrusion Detection System, Neha Rai Khushbu Rai, Khushbu Rai
4. Springer, Machine learning for intrusion detection in MANET: a state-of-the-art survey Lediona Nishani & Marenglen Biba.
5. Springers Effective Classification and Handling of Incoming Data Packets in Mobile Ad Hoc Networks (MANETs) Using Random Forest Ensemble Technique (RF/ET) Anand Nayyar, Bandana Mahapatra.
6. International Journal of Computer Applications (0975 – 8887), An Effective Intrusion Detection System for Routing Attacks in MANET using Machine Learning Technique, Pratik Gite, Sanjay Thakur, Ph.D.
7. Zalte S.S., Ghorpade V.R., Intrusion Detection system for MANET, International Conference for Convergence in Technology.
8. Sujithra L R, Nivethaa V, Pavithra B, Pavithran M, Heterogenous Based Intrusion Detection system in Mobile Ad Hoc Network, International Research Journal of engineering and Technology.
9. Indira N, Establishing a secure routing in MANET using a Hybrid Intrusion Detection System, International Conference on Advanced Computing.
10. Sankaranarayanan.S, Murugabhoopathi.G, Secure Intrusion Detection System in Mobile Ad Hoc Network using RSA Algorithm, Second International Conference on Recent Trends and Challenges in Computational Models.