Black Hole Node Behavioral Analysis based on Hop Count and Timestamp using Machine Learning Algorithm

Irshad Hussain[1], Palagani Vinay[2], P. V. L. Deepthi[3], Raju Mohan[4], Mrs. Y. Adilakshmi[5]

[1]Mohammed Irshad Hussain, Gudlavalleru Engineering College
[2]Palagani Vinay, Gudlavalleru Engineering College
[3]P. V. L. Deepthi, Gudlavalleru Engineering College
[4] Nandeti Raju Mohan, Gudlavalleru Engineering College
[5]Mrs. Y. Adilakshmi: Associate Professor, Dept. of Computer Science and Engineering, Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India.

---***---

Abstract - MANETs being popular for their efficiency and features like Minimal configuration and Immediate Deployment makes themselves an interesting area for the researchers. However, they often get subjected to diverse attacks primarily due their non existing physical topology and no centralized monitorization. Relatively, many researches have been going on to identify the intruder and lower the intensity of attacks on the MANETs. One of them is the development of an IDS (Intrusion Detection System). This paper focuses on two most significant attributes and their contribution in classifying a Black node.

Key Words: **MANETs, Intrusion Detection System, Black Node**

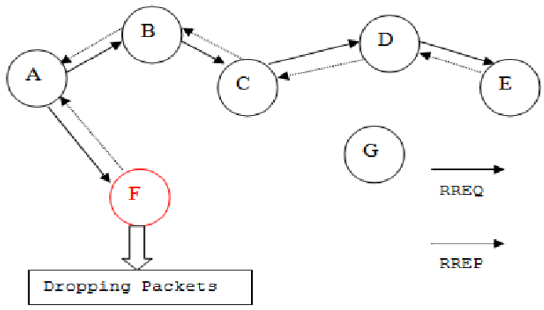
1.INTRODUCTION

An Intrusion Detection System is a software program or application developed to detect the intrusion. IDS can be categorized as signature-based detection system (pattern based) and anomaly-based detection systems (traffic based). However, in order to detect an intruder based on anomaly it is necessary to concentrate on the parameters that have a high significance in detecting the Black Hole Attack. In order to work with the significant attributes, they must be identified first and this can be done by closely observing the Black Hole Attack and discovering the parameters that can provide less positive upon consideration. This paper describes the project carried out to develop an IDS based on the results generated by the simulation of a Black Hole Attack in a node simulator using AODV protocol and discovering two of the most significant attributes that can yield high accuracy when considered.

1.1 BLACK HOLE ATTACK:

The Black holes in the network cause damages:

- Behaves as if it is a Source node by faking the Route Request packet.
- Behaves as if it is a Destination node by faking the Route Reply packet.
- Decreases its hop count value, when it forwards Route Request packet.



A has to send packets to G and sends request to its neighboring nodes for a route. Here F being the black node acts as if there is route from node A to G through it and constitutes Packets Dropping.

1.2 EFEECT OF BLACK HOLE ATTACK ON MANTETs:

- Performance of the network gets effected.
- Packet Dropping increases.
- Packet Delivery Ratio decreases.
- Loss of Information.

The paper is divided into four sections, the first section introduces the work done, followed by the second section representing the Literature Review. Consequently, the third section describing briefly about the Proposed Methodology and Random Forest Classifier. Ultimately the section four for Results and Conclusion.

2. RELATED WORKS:

Many Researches were done in the past to develop a IDS in order to prevent Black Hole Attack. However, these were performed on various Datasets such as:

i. KDD Dataset

ii. DARPA Dataset

iii. CDX Dataset

iv. UNB ISCX 2012

The lack of adequate datasets has led to an anomaly-based IDS. The last experiences suffer with the absence of exact arrangement, examination, and assessment in Mobile Ad-Hoc Network. Be that as it may, it is hard to track down suitable and substantial datasets to assess a versatile system.
the underneath conversation about the applications for interruption identification in MANET utilizing SVM calculation, and existing datasets give different subtleties to concentrate on issues that should be unraveled or amplified. Nevertheless, the data for the training and testing is a very critical issue.
They can be obtained from any of the two ways real traffic or simulated traffic. The real traffic is very costly; the Ad-Hoc system can utilize a conveyed and neighbourhood pruning procedure to choose the sending hub among the sending given Also the past works present that the single confined machine learning calculation would not propose the acknowledged location rate.
The interest is in the most important performance parameters e.g. false negative and false positive to evaluate the selected classifiers. Because of the implemented tests the emphasis will be on choosing the adequacy of the machine learning classifier which accomplished the acknowledged precision rate with the base false negative worth.

Our commitment is to identify indications of intrusion situations by following an improved intrusion detection way and to deal with another dataset.

3. PROPOSED METHODOLOGY

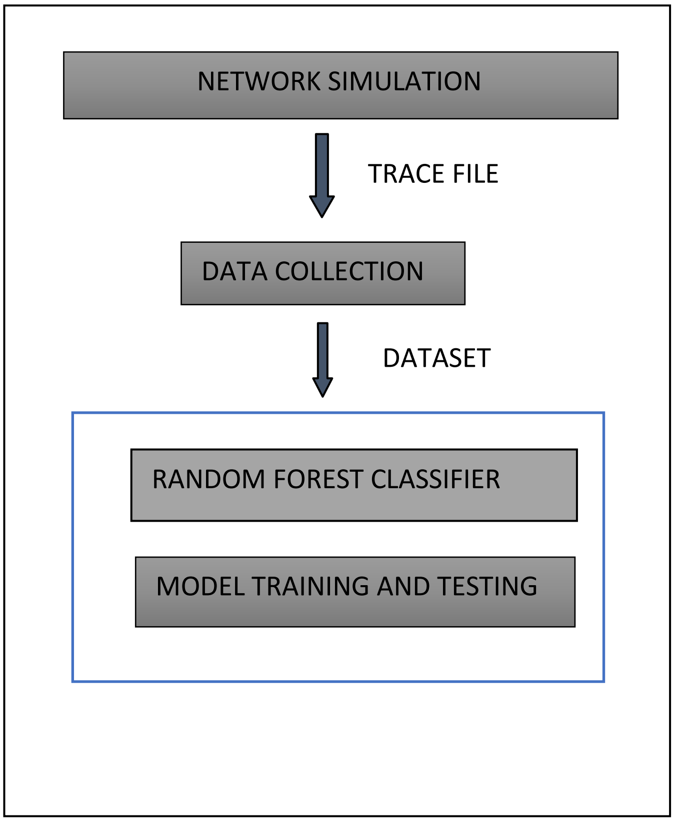
In our methodology we mainly follow three steps:

1.Network Simulation

2.Data Collection

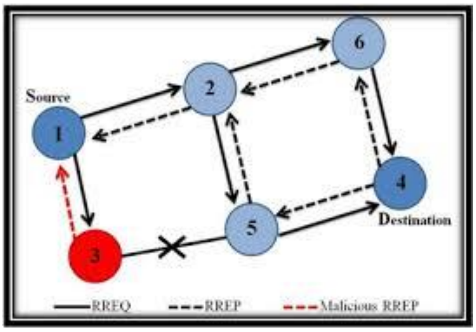
3.Model Training and Testing

When the Black Hole Attack is simulated it generates a trace file which if analyzed, could be used to prepare a labelled dataset which in further can be used to train the data with a machine learning algorithm that would classify between black node and white node. By this the black nodes can be identified and removed from the network.

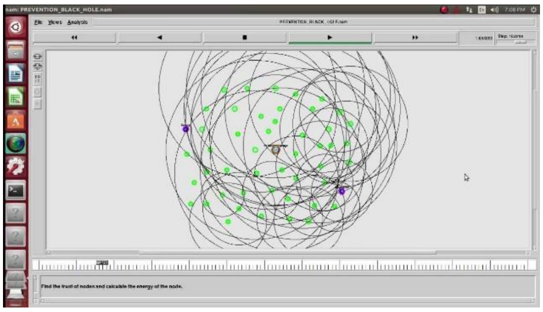


3.1. Network Simulation:

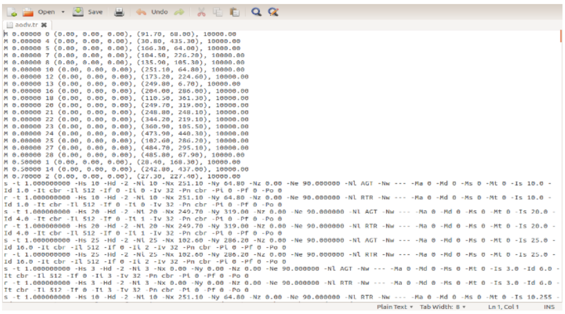
A virtual mobile ad hoc network is stimulated in ns2 tool with 25 nodes and black hole attack is generated.



It is assumed that there are some malicious nodes in the network causing the black hole attack and based on their anomaly behavior they are detected.



The trace file then generated is used to collect the data:

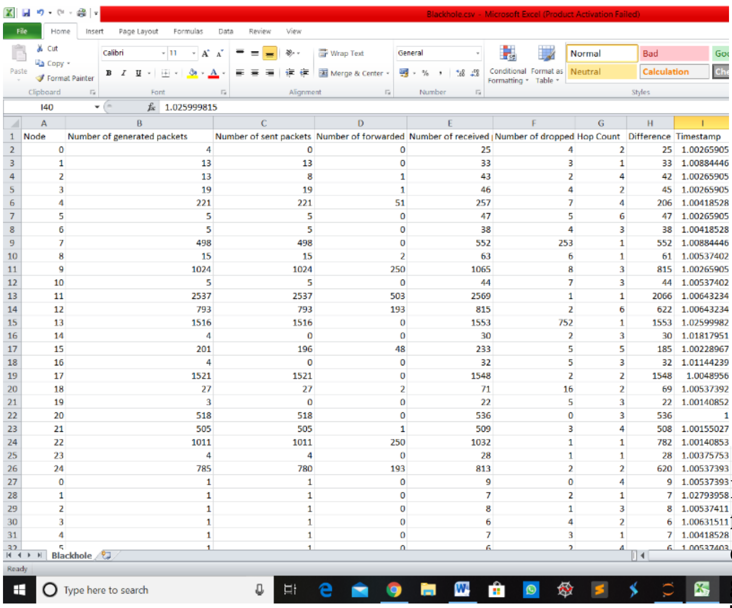


3.2. Data Collection:

The trace file is then observed in detail in order to create a csv file.

Based on the trace file generated by the simulation of a Black Hole attack and with reference to the related works some important attributes having high significance in the attack are extracted and then they are stored in a dataset.

The nodes consisting anomalies in their behavior are identified and considered black holes.



	Node	Number of generated packets	Number of sent packets	Number of forwarded	Number of received	Number of dropped	Hop Count	Difference	Timestamp
2	0	4	0	0	25	4	2	25	1.00265905
3	1	13	13	0	33	3	1	33	1.00884446
4	2	13	8	1	43	2	4	42	1.00265905
5	3	19	19	1	46	4	2	45	1.00265905
6	4	221	221	51	257	7	4	206	1.00418528
7	5	5	5	0	47	5	6	47	1.00265905
8	6	5	5	0	38	4	3	38	1.00418528
9	7	498	498	0	552	253	1	552	1.00884446
10	8	15	15	2	63	6	1	61	1.00537402
11	9	1024	1024	250	1065	8	3	815	1.00265905
12	10	5	5	0	44	7	3	44	1.00537402
13	11	2537	2537	503	2569	1	1	2066	1.00643234
14	12	793	793	193	815	2	6	622	1.00643234
15	13	1516	1516	0	1553	752	1	1553	1.02599982
16	14	4	0	0	30	2	3	30	1.01817951
17	15	201	196	48	233	5	5	185	1.00228967
18	16	4	0	0	32	5	3	32	1.01144239
19	17	1521	1521	0	1548	2	2	1548	1.0048956
20	18	27	27	2	71	16	2	69	1.00537392
21	19	3	0	0	22	5	3	22	1.00140852
22	20	518	518	0	536	0	3	536	1
23	21	505	505	1	509	3	4	508	1.00155027
24	22	1011	1011	250	1032	1	1	782	1.00140853
25	23	4	4	0	28	1	1	28	1.00375753
26	24	785	780	193	813	2	2	620	1.00537393
27	0	1	1	0	9	0	4	9	1.00537393
28	1	1	1	0	7	2	1	7	1.02793958
29	2	1	1	0	8	1	3	8	1.00537411
30	3	1	1	0	6	4	2	6	1.00631511
31	4	1	1	0	7	3	1	7	1.00418528

The dataset CSV file has the following attributes

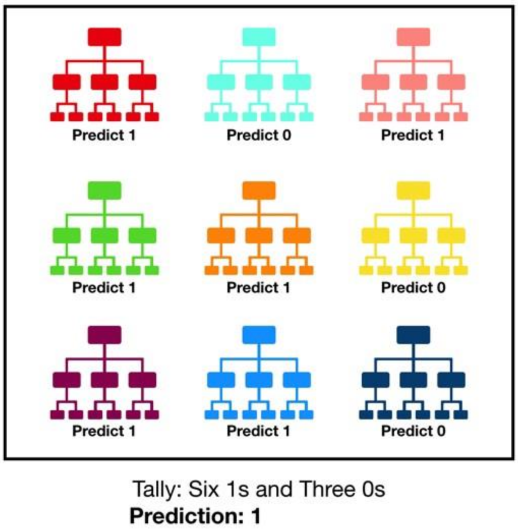
	Traffic data related features:CBR Traffic
V1-V3	Number of sent, received and forwarded CBR data packets NumSentCbrPkt, NumRecvCbrPkt, NumFwdCbrPkt
	Path discovery related features: RREQ
V4	Number of sent packets *NumSentRReqPkt*
V5-V7	Number of received packets: with the same source address as the node *NumRecvSameSrcRReqPkt* with the same destination address as the node *NumRecvSameDstRReqPkt* with the different source and destination address of the node *NumRecvDiffSrcDstRReqPkt*
V8	Number of forwarded packets *NumFwdRReqPkt*
	Path discovery related features:RREP
V9-V10	Number of sent packets : with the same destination address as the node *NumSentSameDstRRepPkt* with the different destination address of the node *NumSentDiffDstRRepPkt*
V11-V12	Number of received packets: with the same source address as the node *NumRecvSameSrcRRepPkt* with the different source address of the node *NumRecvDiffSrcRRepPkt*
V13	Number of forwarded packets *NumFwdRRepPkt*
	Path interruption related features
V14-V16	Number packets of sent, received and forwarded RERR *NumSentRErrPkt, NumRecvRErrPkt, NumFwdRErrPkt*
V17-V18	Number of dropped [RREQ/RERR] packets *NumDropRReqPkt, NumDropRRepPkt*
	AODV protocol specific feature
V19	Average difference at each time slot between destination SN of received RREP packet *AvgDiffDstSeqNum*
V20	Average differences between the magnitude of HC of received RREP packet *AvgDiffDstHopCount*

While preparing the data it is essential to consider parameters which have higher significance so That max noise is reduced. The Hop count value and the Timestamp values are considered to be high significant parameters.

These values are obtained from the analysis of trace file generated by the simulation of black hole attack simulated in the ns2 tool.

While collecting the attributes we have observed that the nodes were showing high anomalies in their hop count and timestamp values as hop count of an individual node is the number of nodes between the itself and receiver. Generally, in most of the cases black nodes project themselves as the nodes having minimal hop count in the network. In addition to this the nodes were having higher Timestamp than the usual.

3.3. Training on Random Forest Classifier:



Random Forest Classifier, the name itself suggests that it a forest i.e., a collection of decision trees that operates as an ensemble. These trees predict values and the value which gets maximum votes is considered as the required result. The basic methodology behind Random Forest classifier is robust

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

A bundle of relatively uncorrelated trees operating as a group will perform any of the individual constituent models.

The low correlation between models is the key. Low correlations form a committee together to produce ensemble predictions making them more accurate than individual predictions

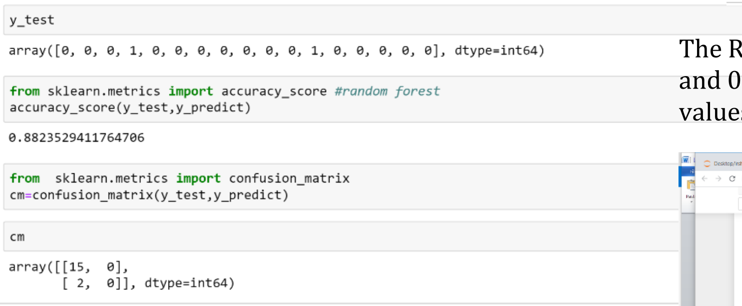
This is because of the group formation of all these trees. Even there is a chance that some trees may individually move in the opposite direction but as a batch the trees all together move in the correct direction. So, the prerequisites for random forest to perform well are:

1. There is a need of some sort of real hint in our features so that the models built using those features do better than random guessing.

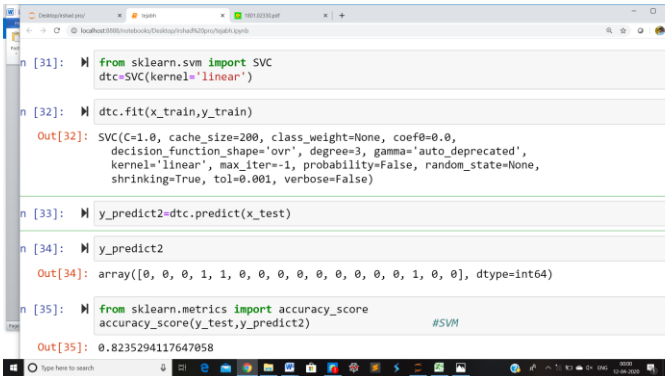
2. The predictions made by the single trees should have less correlations with one another.

4.RESULT AND CONCLUSION

When the model was trained with Random Forest Classifier it was seen that the model scored an accuracy of 0.88

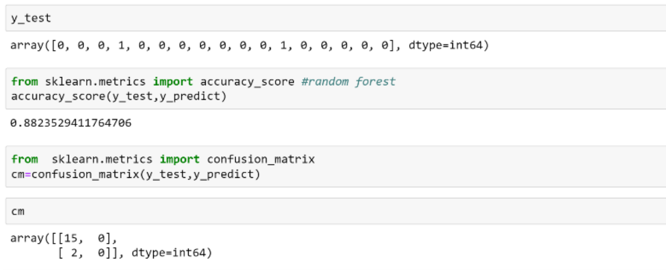


The model was also trained with SVM classifier to see the accuracy score and found that using SVM classifier scored an accuracy of 0.82

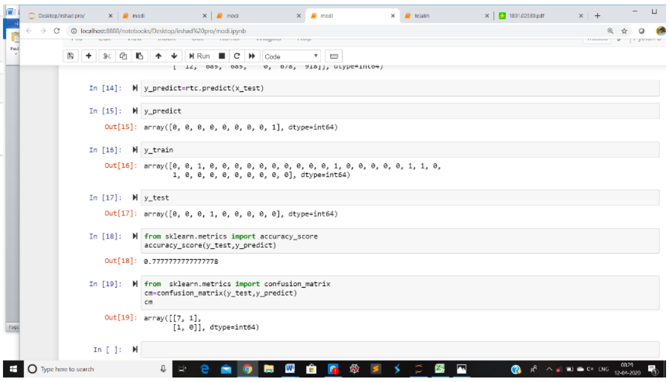


This is because of the Uncorrelated Data Random Forest performed well.

Moreover, the Model was trained with and without considering Hop count and Time Stamp and found that including these parameters increases the accuracy score.



The Random Forest yielded an accuracy score of 0.88 with and 0.77 without including Hop count and Timestamp values.



CONCLUSION

To conclude, we have simulated a mobile ad hoc network and worked on the trace file to observe the anomaly behavior of nodes and prepared a dataset based on the simulation. On Training the dataset with Random Forest Classifier we have found that the accuracy score of the dataset was 0.77 and thus accuracy rose to 0.88 when the timestamp and hop count was included in the dataset. In future some extensions can be done to the project by simulation a network with a greater number of nodes as well as adding some other attributes to the dataset.

REFERENCES

1. The Hundred Page Machine Learning Book -Andry Burkov

2. Abdel-Fattah, F., Dahalin, F., Jusoh, S., 2010. Distributed and cooperative hierarchical intrusion detection on manets. International Journal of Computer Applications 12.

3. Deng, H., Xu, R., Li, J., Zhang, 2006.International Conference on Parallel and Distributed Systems.

4. machine learning for absolute beginners -O.THEOBALD

5. Huang, Y.a., Fan, W., Lee, W., Philip, S.Y., 2003. Cross-feature analysis for detecting ad-hoc routing anomalies, in: Distributed Computing Systems, IEEE. p. 478.

6. Jain, A., Nandakumar, K., ROSS, A., 2005. Score normalization in multimodal biometric systems. Pattern Recognition 38, 2270–2285.

7. Ad hoc networks modified by Jesus Hamilton Ortiz.

8. Understanding Machine Learning: From Theory to Algorithms By Shai Shalev- Shwartz and Shai Ben-David