

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

Irshad Hussain¹, Palagani Vinay², P. V. L. Deepthi³, Raju Mohan⁴, Mrs. Y. Adilakshmi⁵

¹Mohammed Irshad Hussain, Gudlavalleru Engineering College
 ²Palagani Vinay, Gudlavalleru Engineering College
 ³P. V. L. Deepthi, Gudlavalleru Engineering College
 ⁴ Nandeti Raju Mohan, Gudlavalleru Engineering College
 ⁵Mrs. 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.

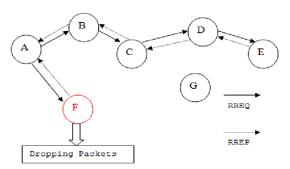
1.1 BLACK HOLE ATTACK:

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.

NETWORK SIMULATION

DATA COLLECTION

RANDOM FOREST CLASSIFIER

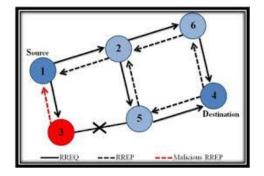
MODEL TRAINING AND TESTING

TRACE FILE

DATASET

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.



International Research Journal of Engineering and Technology (IRJET) e-Is

Volume: 07 Issue: 05 | May 2020

www.irjet.net

e-ISSN: 2395-0056 p-ISSN: 2395-0072



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

	3 1 m	171 de made	AL 14 17 1		~										
		🛎 😽 Oneo 2	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~		14										
		00) (91.70 65	1000 10000 0												
<pre> Boole 1 (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1</pre>	M 0.00000 4 (0.00, 0.00, 0.	00), (30.80, 43	15.30), 10000.	00											
<pre> A mode 0 { 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 +</pre>															
<pre> Borness 1 (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1</pre>															
<pre> Benedic 1 (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1</pre>															
	M 0.00000 12 (0.00, 0.00, 0	.08), (173.20,	224.60), 1000	8.66											
<pre></pre>															
θ e conse a la cará a															
<pre>0 0.0000 01 (0.000, 0.000</pre>															
	M 0.00000 23 (0.00, 0.00, 0	0.08), (368.90,	105.50), 1000	0.00											
n e esco z 10 (zml, 10, ml, 10, ml), 400 (zml, 10, ml), 100 (zml, 10,	M 0.00000 25 (0.00, 0.00, 0	.08), (102.60,	286.20), 1000	0.00											
<pre>6 0.0000 [1 0.000] 0.000 [1 0.000] (1 0.000] 0.000 0.000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.000000</pre>															
1. L. CONSTRUCTION THE CALL LARGE VIEW STATES AND															
10 L D C LLC (C LL C LL C LL C LL C LL C L	M 0.70000 2 (0.00, 0.00, 0.	00), (27,30, 22	(7,40), 10000.	00											
						.000000 -N	L AGT	-Ne	Ma 0 -	-md o	-Ms 0	-nt e	-15 :	10.0 -	
1 C. DORESSEN M. 17 MI 2 MI						.000000 1N	CRIR	- New	na o .	Ma o	-12 0	-ne e	1.12	10.0	
							NL AGT	-Ne	-Ma 0	-Md 0	-Ms 0	- 11	0.15	20.0	
16 4.6 (16.00) 17.2 (17.0) 17.1 (1.1) (1.1	Id 4.0 -It cbr -Il 512 -If	0 -T1 1 -TV 32	-Pn cbr -Pt 0	-Pf 0 -P	0.0										
						0.000000 -1	NL RTP	-Nw	-Ma 0	-Md 0	-Ms (-Mt	0 -15	20.0	
10 45.6 - 11.2 10.0000000 10.1 20 - 17 6 - 11 2 - 12 2 - 193 65 14 0 - 17 6 - 1															
						0.000000 -1	NU AGI	-Nw	-Ma 0	-Md 0	-M5 (- AC	0 -15	25.0	
									- 14 - 10	- Mall of	- 20- 0			20.00	
te che 11 512 -176 0 -111 -174 32 -174 che -141 δ -147 6 -10 0 er 1. 500008006 183 - 144 -2 -141 - 148 x 0,08 196 000 -147 60,0 -146 00,000000 -141 878 -144 - 148 0 -145 8 -145 6 -15 3,0 -176 6,0 Te che -11 513 -177 0 -113 - 174 32 -174 che -174 0 -04 0 - 51 - 1. 16000-1600 -145 10 -141 2 - 141 0 -147 6 -106 0	Id 10.0 -It cbr -Il 512 -If	0 -IL Z -IV 32	Pn cbr ·Pi /	0 - Pf 0 -	Po 0										
r + 1.0000000000 HK 3 HH -2 HK 3 - HK 3.00 - HY 0.00 - HK 0.00 +KK 0.00 +0000000 - NL BTR -Nx	s -t 1.000000000 -Hs 3 -Hd	-2 -Ni 3 -Nx 0.	00 -Ny 0.00 -I	Nz 0.00 -	Ne 90.0000	DO -NÌ AGT	-Net -	Ma D	-Nd 0	-Ms 0	-Mt C	- In	3.0 -:	Id 6.0	
Tt cbr -TL 512 -Tf 0 -TL 3 -TV 32 -Pn cbr -PL 0 -Pf 0 -Po 0 s t 1.000000000 -HS 10 -Hd -2 -NL 10 -NX 251.10 -NY 64.80 -NZ 0.00 -Nc 90.000000 -NL RTR -NwNa 0 -Hd 0 -NS 0 -Nt 0 -TS 10.255															
s t 1.000000000 Hs 10 Hd -2 HL 10 Nx 251.10 Ny 64.00 Nz 0.00 Nc 90.000000 NL RTR Nw Na 0 Hd 0 Hs 0 Nt 0 Is 10.255	r -t 1.000000000 -Hs 3 -Hd	-2 -N1 3 -Nx 0.	.00 Ny 0.00	Nz 0.00 -	No 98.0000	00 -NL RTR	- Nw -	Ma 0	-Nd 0	-Ms 0	-Mt (-Is	3.0 -1	Id G.C	
	IT CDF -11 812 -17 9 -11 3	-1V 32 -Ph CDP	251 10 -NY 64	00 .NT 0	00 No 00	000000 -N	1 0 7 0	- Nov	No. 0	M4 0				10.255	
		· · · · · · · · · · · · · · · · · · ·	Correction only be	100 182 0	100 .80 90										

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.

」り ・(* - ∰ = File Home Insert Page L	ayout Formulas Data	Review View	Blackhole	csv – Microsof	t Excel (Product Activation	on Failed)				44	.	式式式	
🗎 👗 Cut 🛛 Calibri	• 11 • A' A' =	• = 😑 🗞 - 🛛 📅 Wra	ip Text General		Normal		Bad	Go		Predict 1	Predict 0	Predict 1	
aste	- E - & - A - =		rge & Center + 🔫 + % 🔸 🐒		nal Format as Neutral		Calculati						
Clipboard IV	Font la	Alignment	Fi Number	Formattin	g * Table *	St	des		I .				
140 - (*	£ 1.025999815						1			and the second se			
A B	3	С	DE		F G		н	1				1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
Node Number of genera	ted packets Numbe	er of sent packets Numbe	er of forwarded Number of re	ceived Numb	er of dropped Hop Co	ount Dif	ference 1	imestamp					
0	4	0	0	25	4	2		1.0026590					
1	13	13	0	33	3	1		1.0088444	1	ي الم الم			
2	13	8	1	43	2	4		1.0026590			يف يف يف يف يف	***	
3	19 221	19 221	1	46 257	4	2		1.0026590		Predict 1	Predict 1	Predict 0	
4	221	221	51	47	5	4		1.0041852	1 1		. rouide /		1
6	5	5	0	38	4	3		1.0041852	1				
7	498	498	0	552	253	1		1.0088444					
8	15	15	2	63	6	1	61	1.0053740					
9	1024	1024	250	1065	8	3	815	1.0026590					
10	5	5	0	44	7	3		1.0053740					
11	2537	2537	503	2569	1	1		1.0064323					
12	793	793	193	815	2	6		1.0054323					
13	1516	1516	0	1553 30	752	1		1.0259998		د خ خ د			
15	201	196	48	233	2	5		1.0022896		Predict 1	Predict 1	Predict 0	
15	4	0	-10	32	5	3		1.0114423					
17	1521	1521	0	1548	2	2	1548	1.004895					4
18	27	27	2	71	16	2	69	1.0053739				annian a inn	
19	3	0	0	22	5	3	22	1.0014085		lall	y: Six 1s and Thi	ree 0s	
20	518	518	0	536	0	3	536						
21	505	505	1	509	3	4		1.0015502		Pre	diction: 1		
22	1011	1011	250	1032	1	1		1.0014085					
23	4 785	4 780	0 193	28 813	1	2		1.0037575	0				
24	1	1	193	0	0	4						1. 10	
1	1	1	0	7	2	1	7	1.0279395	Kandon	i Forest	Classifier the	name itself sugg	ests that it
2	1	1	0	8	1	3	8	1.0053741	0				,
3	1	1	0	6	4	2	6	1.0063151	forest i.	e., a coll	ection of decis	ion trees that op	perates as a
4	1	1	0	7	3	1	7	1.0041852		e., a com	control of accels	ion a cos anac op	
▶ H Blackhole	1	1	0	6	2	4	4 6	1.0053740	ensemb	le Thes	e trees predict	values and the	value whicl
h buckindle / Ca						0			enseme	nei mes		, values and the	value willer
Type here to search	ch	0 H C	🚖 🧿 🖿 🛙	B 🔒	<u>o</u> 🅸 롣	\$	С	8	gets ma	🏹 ng ung	10 2 (3) INC S 1815 5	ered as the requ	ired result.
									The bas	ic meth	odology behind	d Random Fores	t classifier i

The dataset CSV file has the following attributes

V1-V3	Number of sent, received and forwarded CBR data packets NumSentCbrPkt, NumRecvCbrPkt, NumFwdCbrPk
	Path discovery related features: RREQ
V4	Number of sent packets NumSentRReqPkt
	Number of received packets: with the same source address as the node NumRecvSameSrcRReqPkt
V5-V7	with the same destination address as the node NumRecvSameDstRRegPkt
	with the different source and destination address of the node NumRecvDiffSrcDstRReqPkt
V8	Number of forwarded packets NumFwdRReqPkt
	Path discovery related features:RREP
No 1/10	Number of sent packets : with the same destination address as the node NumSentSameDstRRepPkt
V9-V10	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, NumFwdRErrPka
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:

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

y_test
array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int64)
from sklearn.metrics import accuracy_score #random forest
accuracy_score(y_test,y_predict)
0.8823529411764706
from sklearn.metrics import confusion_matrix
cm=confusion_matrix(y_test,y_predict)
cm
array([[15, 0],
 [2, 0]], dtype=int64)

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

Desktop/inshed pro/	🛪 🥔 trijsteh 🛪 🚺 1001.52330.pdf 🛪 🛉		
← → C ② local	host 8888 in otebooks; Desktop/irshad%20pro/highth.lpynb	0,	\$ 0
n [31]: 🗎	<pre>from sklearn.svm import SVC dtc=SVC(kernel='linear')</pre>		
n [32]: 🕅	<pre>dtc.fit(x_train,y_train)</pre>		
Out[32]:	SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0, declsion_function_shape='ovr', degree=3, gamma*auto_deprecated', kernel='linear', max_iter=-1, probability=False, random_state=None, shrinking=True, tol=0.001, verbose=False)		
n [33]: 🕅	<pre>y_predict2=dtc.predict(x_test)</pre>		
n [34]: 🕅	y_predict2		
Out[34]:	array([0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0], dtype=int64)		
n [35]: 🕅	<pre>from sklearn.metrics import accuracy_score accuracy_score(y_test,y_predict2) #SVM</pre>		
0.001251	0.8235294117647058		

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.

y_test					
array([0, 0,	0, 1, 0, 0, 0,	0, 0, 0, 0, 1,	0, 0, 0, 0, 0]	dtype=int64)	
	.metrics import re(y_test,y_pred		#random forest		
0.8823529411	764706				
	n.metrics impor		rix		
cm					
array([[15, [2,	0], 0]], dtype=int6	4)			

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

€ → 0	localhost8660/noteb	pols/Desktpp/ishadf/i20pro/mod/ipynb	& ± 0 🧑
Pal	B + % Q I	L ← ↓ HRun ■ C → Code ▼ E3 L 14, 085, 055, 0, 0/6, 316]; 0(y)(=11(0+)	
	In [14]: H	y_predict=rtc.predict(x_test)	
	In [15]: H	y_predict	
	Out[15]:	array([0, 0, 0, 0, 0, 0, 0, 1], dtype=int64)	
	In [16]: M	y_train	
	Out[16]:	array([0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int64)	
	In [17]: H	y_test	
	Out[17]:	array([0, 0, 0, 1, 0, 0, 0, 0], dtype=int64)	
	In [18]: M	<pre>from sklearn.metrics import accuracy_score accuracy_score(y_test,y_predict)</pre>	
	Out[18]:	0.77777777777778	
	In [19]: H	<pre>from sklearn.metrics import confusion_matrix cm=confusion_matrix(y_test,y_predict) cm</pre>	
1065	Out[19]:	array([[7, 1], [1, 0]], dtype=int64)	
	In []: H		



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. Crossfeature 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