An Intrusion Detection System in MANET Using Whale Optimization Algorithm (WOA)

Sukriti[1], Neha Batla[2]
[1]M.tech Scholar, [2]A.P.,
[12]Yamuna College of Engineering, Yamunanagar

Abstract-Present day scenario the MANET is the very important tool for the control of network with the help nodes present in it. The formation of this type of node is more dynamic in nature the reason behind it various nodes present in the network. There are various attacks are present in the MANET like Dos attack, probe attack, R2L and U2R. These type of attacks harm the MANET. So for the detection of attack and remove from MANET many techniques are used. The optimization plays an important role in this for minimizing the attacks possibilities. The optimization is Whale Optimization Algorithm is used it provide better accuracy for the attack detection in the MANET. This algorithm is applied on the SVM (support vector machine) which is a machine learning approach. The data set is taken by the WOA is NSL-KDD data set. It is a standard data set which is available easily. The experiment is done on this type of data set NSL-KDD and results are analyzed with the other optimization method in the end. The comparison between WOA and GA results will be done in the end of the work. There are three parameters mainly used for the comparison and observation. These parameters are accuracy, sensitivity and specificity.

Keywords- WOA, MANET, GA etc.

I. INTRODUCTION

The MANET system have the traditional wiring approach. It does not have fixed infrastructure. Most of the wired IDSs depend on the real time visitors parse, calculation, filter out it display the traffic of switches, routers and gateways. The traffic is less provide the opportunity to the IDS applicable on MANET system. Each and every node can simplest use the partial and localized sports. Some of traits are present in the MANET system. Furthermore, in MANETs, it is very hard for IDSs to inform the validity of a few operations. For instance, the motive that one node sends out false routing record due to the fact this node is compromised, or the link is damaged because of the physical motion of the node.

The IDS system analyzes the traffic of the network and inform to the network administrator. The IDS system can block the user or source IP address from accessing the network. The main purpose of the IDS system to detect the behaviour of network and misuse of the information.

In this figure the history of intrusion detection system is shown. The NIDS which is community based and HIDS which is primarily based detection structure are shown. The IDS which is basically based on the precise signature of recognised attacks as like the antivirus in the software which provide protection against the threats. The network intrusion detection system is located at the points in the network to screen visitor which provide community to all the factors. The HIDS system run on the intrusion on person host or device to the network. A HIDS video display units the inbound and outbound packets from the tool only and could alert the person or administrator of suspicious activity is detected.

- It is used wireless link. The data is send in to open condition by the mobile node which is not easily use for the interfacing. In case of wired network the attacker not required any help from the network.
- The other concern is about the MANET not have any equivalent. For the ad-hoc network the good protocol are used.
- All the nodes are dependent to each other so they have not trust and can be easily captured by attackers.
- The system does not have centralized authority.
- Many parameters like as limited bandwidth, battery power or slower links all these parameters do not need any wire

II. NSL-KDD DATA SET

Generally the KDD 99 is used for the data set from (1999). It is prepared by the stolfo[5] and the data built base captured in the DARPA 98. The DARPA system have the 4 GB of of compressed raw (binary) tcp dump data of 7 weeks of network traffic, the system can be processed in to five million connection record where the range is generally 100 bytes. Around 2 million connection records are contains in the two weeks of test data. There are approximately 4,900,000 single connection vectors are contains in KDD training dataset. Each vector contains 41 features and is labelled as either normal or attack. There are 4 categories of attacks in this dataset:

1. *Denial of service attack (DoS):* this type attack effect the memory directly. The memory sources are too busy to handle the requests. The memory sources are too busy for the processing of the machine algorithm.
2. *User to root attack (U2R):* in this type of attack the attackers hack the account of the normal user. So the

system is totally unsafe to get root access to the system. It is a class of exploit.

3. *Remote to Local Attack (R2L):* when attacker efforts some unsafe to gain local access as a user of that machine and also has the ability send the packets over the network to the machine called the remote to local attack. It also not have account on the machine.
4. *Probing attack:* In this type of attack, attacker get round or find a way round of the security controls of the system after getting the information about a network of computer.

The KDDCup training set has 494020 records and test set has 311029 records. For the improvement of the detection rate the attacks groups are categorized and similar attacks present in the category. There are 24 attack type in training set and 38 attacks type in data set. 14 attacks are novel attacks. Some major attacks in the dataset are: namely, Denial of Service (Dos), Probing (Probe), Remote to Local (R2L) and User to Root (U2R).

III. WHALE OPTIMIZATION ALGORITHM

Whale optimisation algorithm (WOA) was published first in 2016 [2] and it formulated the foraging behaviour of humpback whales. The humpback whales feed themselves by consuming small fishes near the water upper surface. So they make a spiral or 9 like structure by bubbles around their prey. During foraging process they go down to 12 meter and make bubbles in circle to confuse the fishes and then swim up toward the surface as shown in figure 3.1. This foraging behavior is called bubble net feeding method. The movement of whale is classified in two ways, one in which whale goes down, make bubbles and then goes up and second one includes different stages: coral loop, lob tail, and capture loop. The encircling prey is to reach to optimal position which is position of fishes. To get this thing done whale must use some random seed (random search space) so that based on the optimal position can be obtained. since at this stage optimal position is not known so it is assumed that current position is the optimal one. Whale is a social animal when it gets the optimal position, other whales are also attracted towards that and updated their positions using equation 3.1.

$$\vec{D} = \left|\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)\right|$$

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - \vec{D}.\vec{A}$$

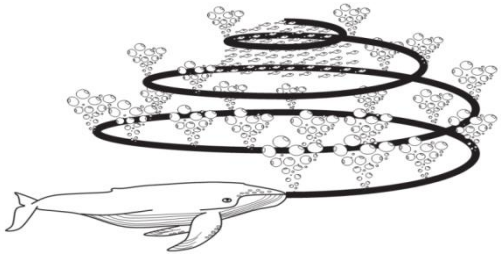

Fig.1: Bubble-net feeding behaviour of humpback whales [2]

$\vec{C}$ & $\vec{A}$ are coefficients, $\overrightarrow{X^*}$ is the position vector of best position obtained so far and $\vec{X}$ is the position vector. The A & C are calculated as:

$$\vec{A} = 2.\vec{a}\vec{r} - \vec{a}$$

$$\vec{C} = 2.\vec{r}$$

where $\vec{a}$ decreased from 2 to 0. The searching location dimension of whale depends upon the number of tuning variables.

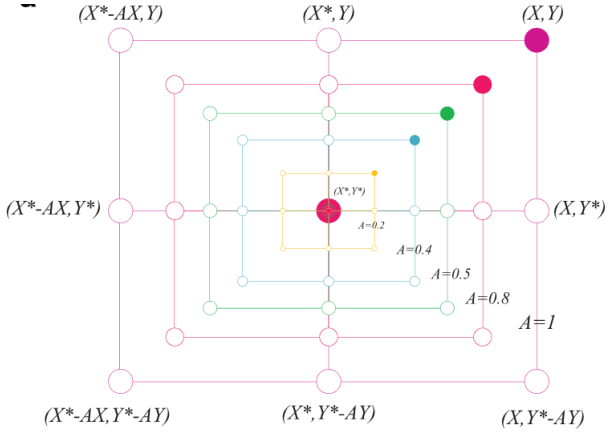

Fig.2: Shrinking Encircling mechanism in bubble net attack approach [2]

The bubble net attacking method [2] in which whale reaches near to optimal position has two approaches: shrinking the encircling mechanism, reducing the spiral path. In shrinking encircling approach the a is reduced in equation 3.3 and corresponding A will also change. The random value of A is chosen in between [-1,1] and new position can be anywhere in between the current position and current best position by equation 3.2. Figure 3.2 shows the possible positions of the whale towards optimal position (X^*,Y^*) in 2D space.

In spiral updating position approach, to move the whale in helix manner, a spiral equation is designed which makes whale to move in spiral fashion towards prey.

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) + \overrightarrow{D'}.e^{bl}.\cos(2\pi l)$$

where D' is the position difference of current whale position and prey position. To reach to an optimal position both approaches is to be used which is decided by a probability factor p. If p<0.5 then shrinking encircling approach will be active else updating spiral position as whale takes 12 meter dive downwards and then generate bubble net, so to reach to prey it has to reduce the circle as well as spiral path as shown in figure 2. The whale optimisation method is considered as global optimisation as it satisfies the theoretical four global conditions: it can switch easily between

exploitation/exploration phase, it shares its current best location with nearby whales, adaptive variation of *A* allows the smooth transition between exploration-exploitation phase and it has only two internal parameters (*A & C*) to tune for best results unlike evolutionary optimisation methods.

IV. PROPOSED WORK

We have divided our work in six sub cases which are

1. Take the NSL-KDD data set and consider the training and test data set from the master one data set.
2. After the step one then applied WOA for the feature reduction. The maine purpose is identified the attack with higher accuracy.
3. Apply training and testing data in to the SVM classifier with the reduce feature
4. Test data using SVM trained model and check accuracy of predicted labels
5. Compare results with GA reduced feature to proposed method

Feature reduction using WOA

The NSL-KDD have the large number of feature in case of training and testing the the data the features are 41. Before using the data into machine learning (ML) model, we need ot preprocess it as data is in raw format and not ready to use as it is. Every ML model understand the langaugae of numerics only. Data has strings in the features, we need to first convert them to analytics. Many attributes has maximum number of zeros which don't contribute in classification and biase the network training. So we select them programmetically and removed from the features set. Some features have high numeric values and some of them have comparatively small values. This large difference also bias the machine learning model. So we normalised them as:

$$normalised\ attribute = \sum_{i=1}^{n} \frac{f_i - \min(f)}{\max(f) - \min(f)}$$

Where 'n' is the total number of samples in an attribute, f_i is the sample value of i^{th} attribute.

After removing the columns with 50% of samples with zero values, we only left with 16 attributes in the data. These all attributes still not contribute in the accuracy but now we don't know which set of attributes should be chosen. For that purpose we opted the novel optimization algorithm based on Wolf's hunting behavior. This optimization is known as Wolf optimization algorithm(WOA) and discussed in previosu chapter. When a large data provided by using the more number of feature the predictive model take lot of times in making traing model. After that more time is consume in the testing and classification. MANET have some applications in the tough catagory so for the more accuracy prediction and reduced the time we used an optimization technique.

The WOA is an iterative type of algorithm which can maximize and minimize any type of objective function. The principle of Whale optimization algorithm explained in the previous chapter. The algorithmis work on the principle of whale catching the food. It detects the food position with respect to the whale movement. The whale position can maximize and minimize with respect to the food location. On the basis of WOA the attack detection process accuracy will be incresed with respect to the attack. The attack can detect in less time in case of WOA.

Whale optimization algorithm and fetaure selection are two isolated algorithgms but these work in a closed loop scenario. A blog diagram representing their communication is shown in figure 4.1.

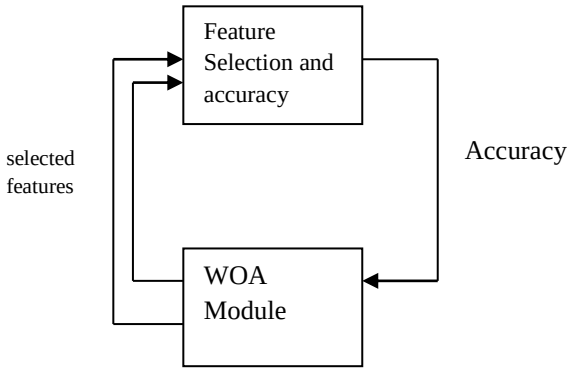

Fig. 3: Relation between feature selection/ML training and WOA optimisation

Both module function in equilibrium, WOA gives the input as binatry matrix to Ml module and gets the accuracy in its input from the ML module. This binary matrix is the set of features which must be included. The '1' in the matrix representes the attribute selected and '0' represents that this feature is not selected. Ml module calculates the accuracy for this set of selected features by training and testing the SVM model. This accuracy is fedback to WOA module which on the basis of it changes the feature set..

V. RESULTS AND DISCUSSUION

In our work we have proposed a comprehensive study on the application WOA (Whale Optimization Algorithm)optimization algorithm for feature reduction for making a better intrusion detection system (IDS).

1. Make availability of NSL KDD dataset and extacting training and testing data from parent dataset.
2. Apply WOA to reduce no. of features which can identify type of attack with higher accuracy.
3. With reduce features, apply training and testing dataset in SVM classifier

4. Test data using SVM trained model and check accuracy of predicted labels
5. Compare results with GA reduced feature to proposed method

Case-1 Denial of Service (DOS) Attack

In NSL-KDD dataset DOS attack is further categorised in six subtypes which are back, land, Neptune, smurf, pod and teardrop. We have compared performance of WOA based feature reduction with GA based feature reduction by using following performance evaluation parameters.

1. Accuracy
2. Precision
3. Recall
4. Specificity

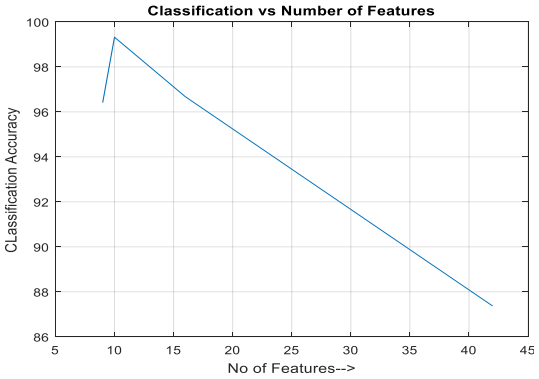

Fig.4: Number of features and its classification

The figure 4 shows the curve between number of classification accuracy and numbers of features applied to the data set. The accuracy obtained during the optimization more than the other method. The next observation is done on the comparison purpose for the GA and WOA.

The above figure shows the comparison across two methods like GA and WOA. The comparison is done on the basis of accuracy, sensitivity and specificity. The result we obtained that WOA provide the better accuracy then the GA. The other two parameters are the equal results but accuracy is more in case of WOA.

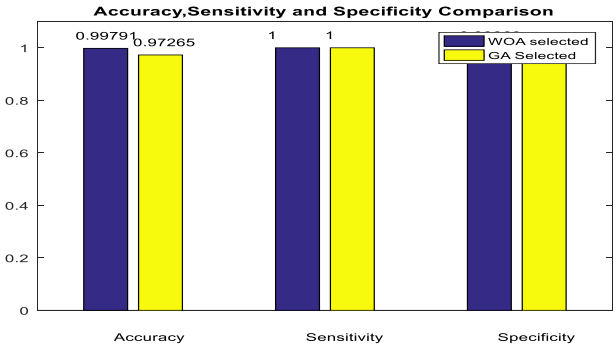

Fig.5: Accuracy, sensitivity and specificity comparison of GA and WOA based method

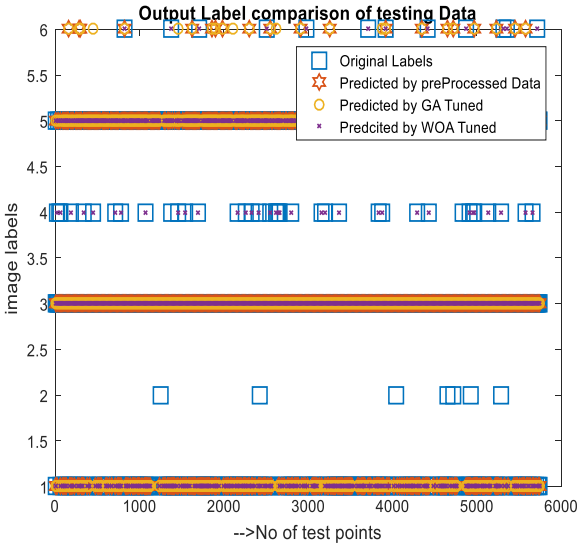

Fig.6 output label comparison of GA and WOA

Figure 5 shows the number of feature selected in the each step and their accuracy with respect to these feature structure. The accuracy which is obtained by the WOA is greater than the accuracy obtained during the GA tuned.

It is observed that proposed method has more accuracy, precision, recall, F measure, sensitivity and specificity for DOS type of attack. For some attack such as pod type DOS attack proposed method gives non-zero value as against standard GA method.

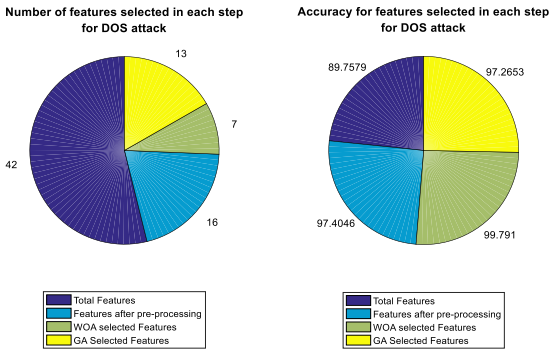

Fig.7 Feature and accuracy of each step in DOS attack.

Case-2 Probe Attack

The figure 8 shows the fitness value and generation of the probe attack. The best fitness and mean fitness function shown by the dotted line or different colors line as shown in the figure. The curve 9 shows the accuracy of the WOA optimization curve. More accuracy is obtained than the GA

case. Approximately 0.99 accuracy analyze in the WOA network.

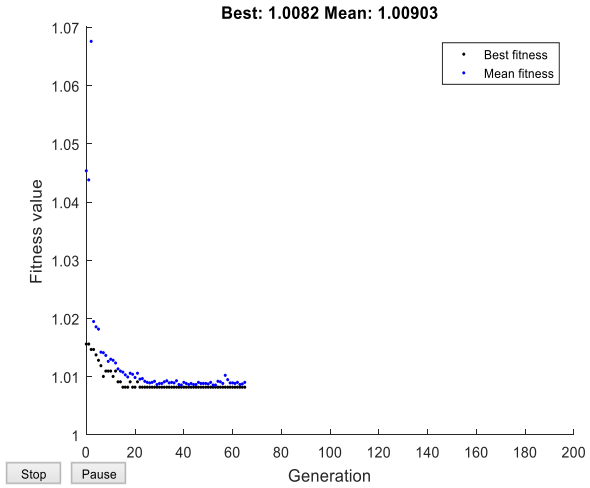

Fig.8: fitness value and generation curve for the best fitness function

The curve 10 shown in the figure accuracy classification and their number of features. As the number of features increased the accuracy is also increase. After the 28 features the accuracy is stable

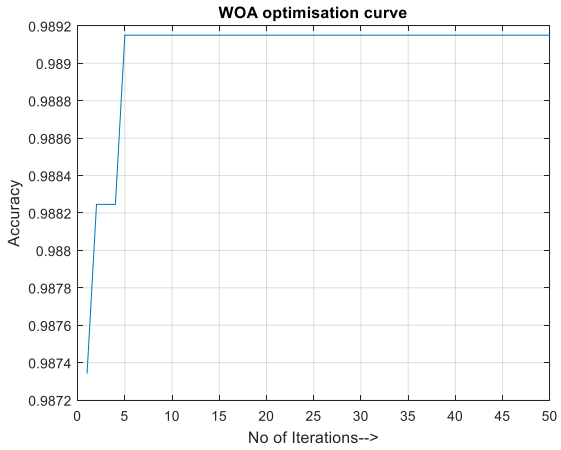

Fig.9: WOA accuracy optimization curve in Probe attack

.

Figure 11 shows the output label data with respect to the WOA and GA. Different parameters are considered in case of label output.

Figure 12 shows the comparison of accuracy with the genetic algorithm. In case of WOA the accuracy results more than the GA.

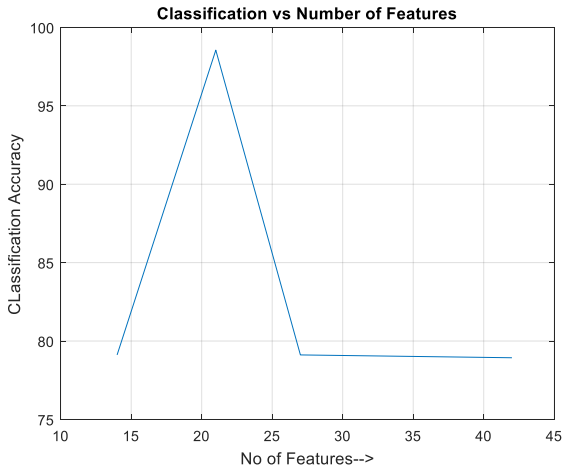

Fig.10: Classification accuracy versus number of features in Probe attack

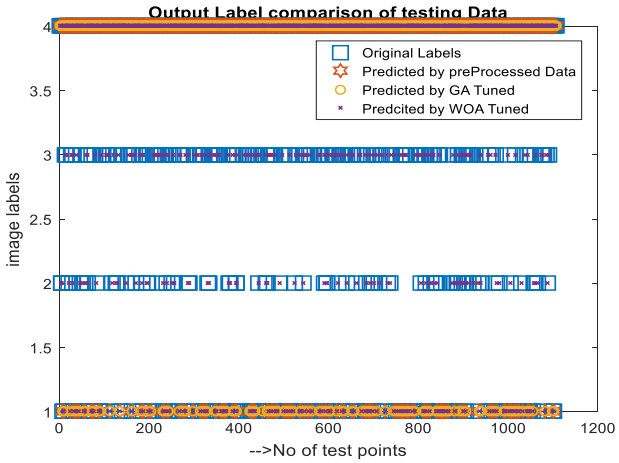

Fig.11: output label data comparison with testing data in Probe attack

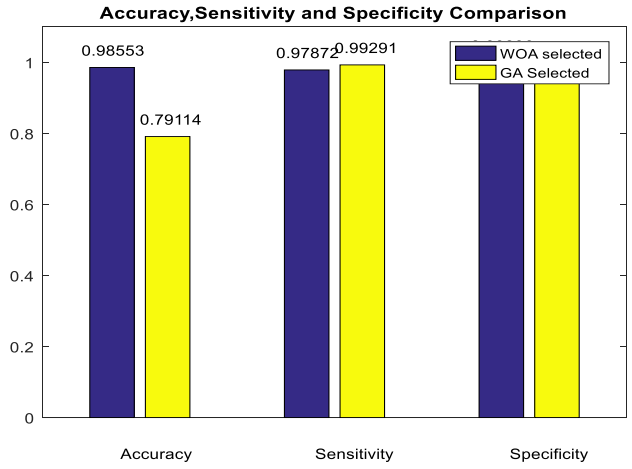

Fig.12: comparison of accuracy, sensitivity and specificity in Probe attack

Figure 13 gives the features selected in each step and their accuracy in case of probe attack. The results are good in case of WOA selected features nearby 98.5533. It is observed from mean values of all parameters of all four type of attack for proposed and standard method, proposed method gives better results for all four types of attacks. Hence we conclude that our proposed method works better in IDS for MANET.

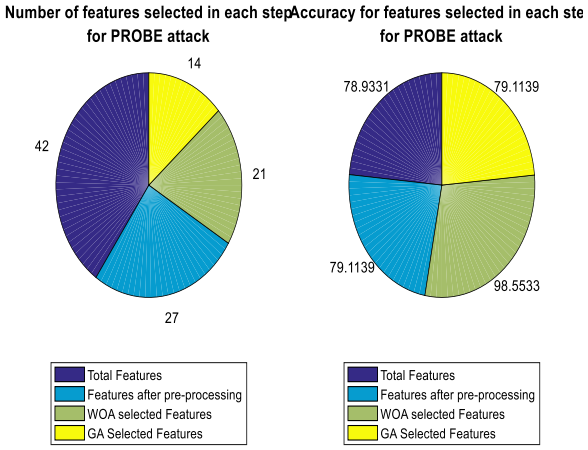

Fig.13: overall analyses with GA and WOA in Probe attack

VI. CONCLUSION

Present scenario the demand of the MANET system is increasing day by day. The Mobile Ad-hoc Network used for different applications like military surveillance and some other important applications. The MANET required the fewer infrastructures so they can easily placed in the area with help of the node. The nodes can easily communicate with the other node which is available in the wide range. There are many types of attack but in our work two attacks are analysed DOS attack and probe attack. Therefore we take NSL-KDD data set and analyze the different cases of attack detection on this data. The optimization process is used for the detection of data set. Whale Optimization Algorithm is considered in this case. Then find out the three parameters like accuracy, sensitivity and specificity of the feature reduction data set of NSL-KDD. A machine learning approach is used for the process of attack detection. Creating a training model using SVM (Support Vector Machine) classifier and method is validated by checking testing data using trained model. Then compare the results with the genetic algorithm. The results shows the better accuracy in case of WOA than the GA.

VII. REFERENCES

[1]. Z. Ullah, M. S. Khan, I. Ahmed, N. Javaid and M. I. Khan, "Fuzzy-Based Trust Model for Detection of Selfish Nodes in MANETs," *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, Crans-Montana, 2016, pp. 965-972.

[2]. M. A. Abdelshafy and P. J. B. King, "Dynamic source routing under attacks," *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*, Munich, 2015, pp. 174-180.

[3]. C. Alocious, H. Xiao and B. Christianson, "Analysis of DoS attacks at MAC Layer in mobile adhoc networks," *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Dubrovnik, 2015, pp. 811-816.

[4]. A. Quyoom, R. Ali, D. N. Gouttam and H. Sharma, "A novel mechanism of detection of denial of service attack (DoS) in VANET using Malicious and Irrelevant Packet Detection Algorithm (MIPDA)," *International Conference on Computing, Communication & Automation*, Noida, 2015, pp. 414-419.

[5]. A. M. Shabut, K. P. Dahal, S. K. Bista and I. U. Awan, "Recommendation Based Trust Model with an Effective Defence Scheme for MANETs," in *IEEE Transactions on Mobile Computing*, vol. 14, no. 10, pp. 2101-2115, Oct. 1 2015.

[6]. A. Menaka Pushpa and K. Kathiravan, "Resilient PUMA (Protocol for Unified Multicasting through Announcement) against internal attacks in Mobile Ad hoc Networks," *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Mysore, 2013, pp. 1906-1912.

[7]. M. A. Abdelshafy and P. J. B. King, "Analysis of security attacks on AODV routing," *8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)*, London, 2013, pp. 290-295.

[8]. A. M. Kurkure and B. Chaudhari, "Analysing credit based ARAN to detect selfish nodes in MANET," *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, Unnao, 2014, pp. 1-5.

[9]. S. Biswas, P. Dey and S. Neogy, "Trusted checkpointing based on ant colony optimization in MANET," *2012 Third International Conference on Emerging Applications of Information Technology*, Kolkata, 2012, pp. 433-438.

[10]. D. Das, K. Majumder and A. Dasgupta, "A game-theory based secure routing mechanism in mobile ad hoc network," *2016 International Conference on Computing, Communication and Automation (ICCCA)*, Noida, 2016, pp. 437-442.

[11]. T. Poongothai and K. Duraiswamy, "Intrusion detection in mobile AdHoc networks using machine learning approach," *International Conference on Information Communication and Embedded Systems (ICICES2014)*, Chennai, 2014, pp. 1-5.

[12]. D. A. Varma and M. Narayanan, "Identifying malicious nodes in Mobile Ad-Hoc Networks using polynomial reduction algorithm," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, 2016, pp. 1179-1184.

[13]. Bandana Mahapatraa and Prof.(Dr) Srikanta Patnaik, "Self Adaptive Intrusion Detection Technique Using Data Mining concept in an Ad-Hoc Network," 2nd International Conference on Intelligent Computing, Communication & Convergence(ICCC-2016)

[14]. Manjula C. Belavagi and BalachandraMuniyal, "Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection," 12th International Multi-Conference on Information Processing-2016 (IMCIP-2016).

[15]. PreetiAggarwala and Sudhir Kumar Sharmab, "Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection," 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).
[16]. Ciza Thomas, Vishwas Sharma and N. Balakrishnan, "Usefulness of DARPA Dataset for Intrusion Detection System Evaluation
[17]. P.Natesan and P.Balasubramanie, "Multi Stage Filter Using Enhanced Adaboost for Network Intrusion Detection," International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.3, May 2012
[18]. M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, 2009, pp. 1-6.
[19]. Seyedali Mirjalili, Andrew Lewis, "The Whale Optimization Algorithm," Advances in Engineering Software, Volume 95, 2016, Pages 51-67, ISSN 0965-9978