Intrusion Detection System using Nature-inspired Metaheuristic algorithms

COMPUTER SCIENCE, ENGINEERING DEPARTMENT,

BENNETT UNIVERSITY, GREATER NOIDA,

UP, INDIA

Hemanth Pedamallu, Sai Sudhamsa Potru, Praneeth Balem, Rishith Nmula

**Abstract**— The IDS has uncertain behaviour on the traffic inside the network because of its large computations. As it's needed to explore the large amount of features, the system complexity also increases. There is a very good chance that the features which are irrelevant and noisy can affect the efficiency of the system. A faster method is proposed for IDS by training machine learning models using NSL-KDD dataset. Many features in the NSL-KDD dataset are not useful to predict an attack. The noisy data has been removed by using Random Forest Classifier. Moreover, the selected features are used to train a feed-forward neural network with various meta-heuristic nature inspired algorithms, which gives good predictions on the test data. The proposed method can train the network to give an accuracy, detection rate and false alarm rate of 95.8%,97.5% and 2.8%. respectively. This method is efficient at identifying attacks in network traffic.

**Index Terms**—Feature selection, PSO algorithm, IDS(Intrusion Detection System), Random forest,

—————————— ◆ ——————————

# 1 Introduction

IDS is the main tool that looks after network traffic for malicious behaviour and triggers warning messages for doubtful activities. It’s an application that scans a system or a network for effective harmful activities. For network or system traffic because of its random behaviour the intrusions make the system impulsive [1]. The integrity, confidentiality, availability and security principles are decreased due to the intrusion and other attacks like fooling, cyber-attacks, traffic analysis and few other harmful exposures are the major reasons for the requirement of an IDS [2]. The signature and anomaly based are the 2 main categories for IDS. The ‘signature type systems’ are also mentioned as “misuse based” IDSs. It compares the signatures of the recognized mean activities and when a match is found it gives alert. Thus, these types of systems are able to detect already known attacks with less false alarm rate [3]. The anomaly type systems are capable of encountering new attack types. This detects the pattern, if the system turns from the general pattern then it sends an alert. This has a high FPR(false positive rate) [4]. The IDS is again classified into 2 types. The first type is the host based system(HIDS) and the second type is network based system(NIDS). The HIDS takes care of the host computers and sends out alerts. But incase of NIDS, it screens the total traffic that passes through the network. If any suspicious activity is similar to already experienced attacks inside the network, an alert message initiates and it will be sent to the network administrator to take the proper action. As the features increase, so the system complexity increases. So, it becomes difficult for the detection system to study the data. It is essential to select useful and important features for the intrusion detection in the field of system security. Selection of significant features is a difficult task because the dataset is composed of many related, unrelated and wasteful attributes, which leads to an increase of the complexity of computation for the detection of intrusion [5,6]. The primary goal of (FS)feature selection is to enhance the effectiveness and efficiency of the detection system. This performed research includes the top features selection to increase the IDSs performance that are having less computational complexity due to the prediction on a set of reduced attributes. Thus, feature selection is done to increase the efficiency of the system and reduce the no. of attributes that are irrelevant before the data pre-processing. Feature selection is observed on the NSL-KDD dataset(it is the updated version of KDD cup 99 dataset) [7], and the implemented method resulted in high accuracy in intrusion detection and also reduced the false alarm rate because of a limited number of features.

# 2 Literature review

Pranav *et al* [8] conveyed FS algorithm plus classifier

by applying the SVM . The paper also amended FS's perusal and classified the intrusion detection program. And also, to illustrate the study call, smooth computing techniques are used in intrusion detection. amir[9] used a PSO algo. performing the FS & PCA for function conversion. The abstract methodology was started using the support vector machine algorithm on KDDCup99 dataset for intrusion observation.The proposed work is extended more with the use of neural lattice by using the NSLKDD dataset. Franco [21] submitted a classification of 17 data et features using the discriminant Fisher rate algorithm. Ian[10] showed a regression tree by using the KDD Cup 99 dataset using a Bayesian network and FS classification for performance success and detection accuracy in the IDS. Zhang introduced many processors for genetic algorithms ( GA) with a rough set for classifying the rules . NSL-KDD and KDD Cup 99 dataset have been used for detailed research for the complete report. The paper defines classifiers, with , single, multiple and ensemble classifiers.. Bharathi [12] used particle swarm optimization algo. is applied for tuning parameters for designing network intrusion detection & multiple criteria linear programming model for making the division. Xue suggested an optimization of the auto adaptive particle swarm to multiple attributes to reduce the dimensions. There are overall 12 datasets performed to show the achievement of the algo. . Bostani used feature selection to get good results from the consisting algo’s using BGS binary gravitational search algorithms including mutual information. Support Vector Machine was applied for reducing attributes as well as NN was applied to find the selected features importance by using DARPA dataset [13].

# 3 Data Resources

In this paper NSL-KDD dataset was used to perform our study. NSL-KDD is a modified version of KDD Cup 99 dataset. This dataset consists of 42 attributes, the 42nd attribute determines the class. In this dataset there are 5 different classes, in that there are 4 attack groups( DoS,

U2R, R2L, Probe) and one is normal. The brief information of each attack class is here

***Denial of Service(DoS):*** an occurrence of disturbance in an accredited user's access to use a network of computer, generally caused by one with harmful intent.

*Ex:* teardrop, smurf, SYN flooding and neptune

***User to Root(U2R):*** unaccredited access to local super

users.

*Ex:* buffer-overflow, rootkit,spy and SQL attacks.

***R2L:*** Type of attack conducted to remotely inappropriately access a specific network address.

*Ex:* Imap, werezmaster, multihope and spy.

***Probe:***It is a type of attack that is consciously to notice and report a "fingerprint" observable in the report..

*Ex:* satan, port-scan, ping-sweep and nmap

***Basic:*** These features include separate control over the TCP transmission to the protocol connections. The no. of features that belong to this category is 9. The detailed view of these attributes is in Table 1.

***Domain Knowledge*:** These are features inside.Total attributes related to this category is 13 .The detailed view of these attributes is in Table 2.

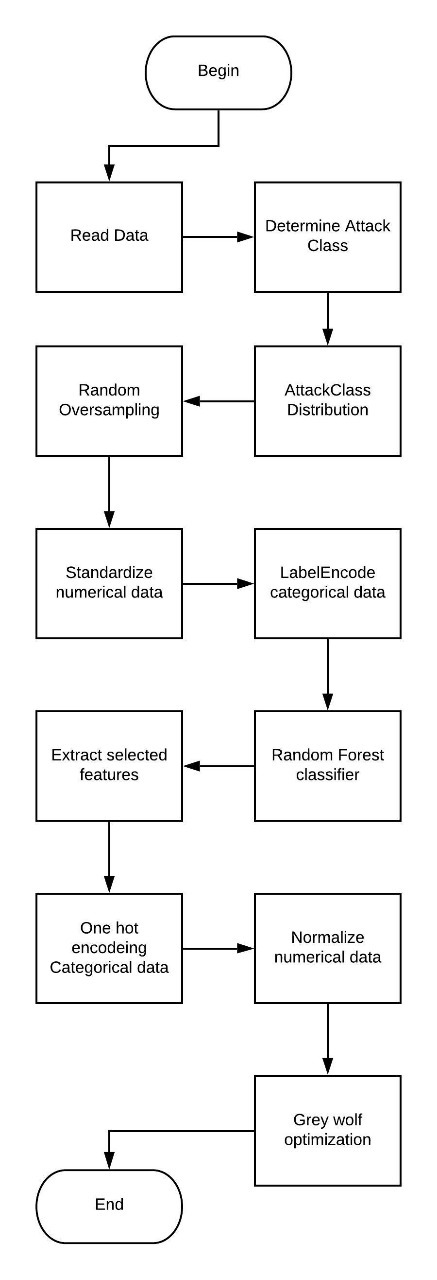
***Traffic:*** This category includes the features counted using 2-s length slots. The no. of features pertaining to the category is 9.The detailed view of these attributes is in Table 3.

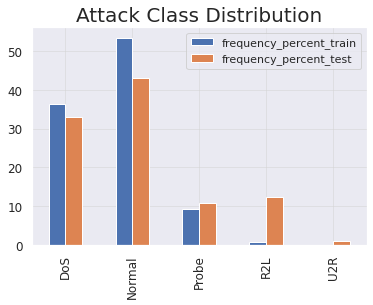
***Host:*** This category contains features that will be represented to test attack duration for > than 2 sec. No. of features in this category is 10.The detailed view of these attributes is in Table 4.

# 4 Methodology

## 4.1 Data preprocessing

As stated in section 3(Data Resources) all the records in the dataset are grouped together and mapped to form 4 attack classes and one normal class. After that, MinMax scaler is applied to all numerical attributes to scale them between 0 & 1. The categorical attributes, including the class attribute are encoded using LabelEncoder. *Figure 2* shows the distribution of the training and testing data, with respect to the attack class. As it is evident from the graph, the number of records in Normal and DoS class are far greater than the number of records in Probe, R2L and U2R. If the machine learning model is trained on this data, there may be a bias towards Normal and DoS class. To avoid this, Random Over Sampler[22] is applied on training dataset to create equal records of each attack class in the dataset. Nilesh et al[14] conducted performance study on different ML classifiers like Naive Bayes(NB), SVM, k Nearest Neighbour, Random forest classifier(RF) and many others for feature selection. The proposed method used RF and selected 10 most significant features. In this analysis, RF is used to derive ten attributes(columns) from 41 attributes. Label Encoder creates a false value hierarchy(greater than, less than relationship) among the categorical attributes. To avoid this, the categorical attributes are re-encoded with OneHot encoder before passing as input to the feed forward neural network for improving its performance.

****

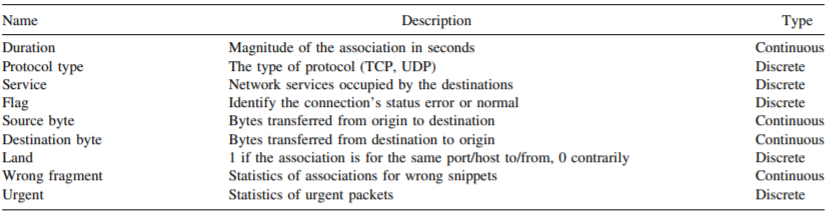


**Figure 2**

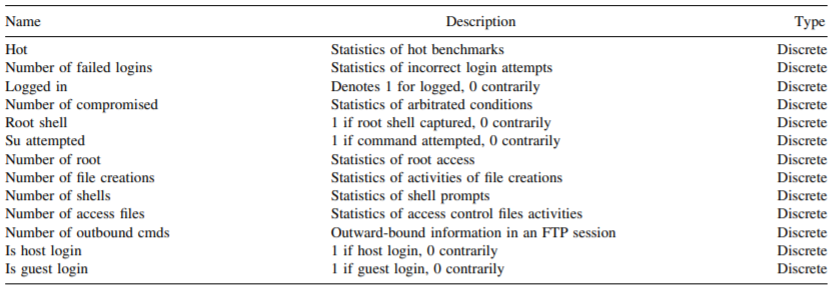
***Figure 1***

Figure 1 shows the general flowchart for the methodology adopted in this research work. The detailed description for each module is elaborated in the following sections.

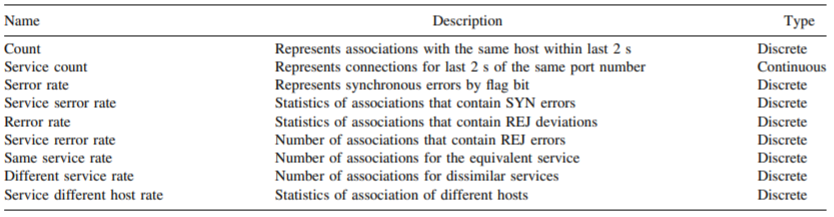
**TABLE 1:** Description of Basic label attributes.



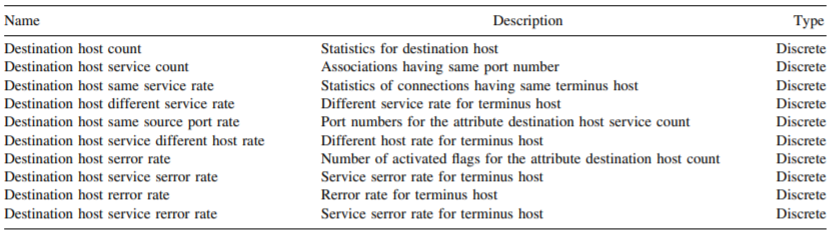
**TABLE 2:** Description of Domain Knowledge label attributes.



**TABLE 3:** Description of Traffic label attributes.



**TABLE 4:** Description of Host label attributes.



## 4.2 Machine learning classifiers

As stated in section 4.1(Data preprocessing) Random forest (RF) classifier is used in this paper for analysing the training part and testing part of data of NSL-KDD dataset.[

c]Random Forest classifier(RF): RF is a ML algorithm, this comprises all the properties of Decision tree(DT) and ensemble learning [15], [16]. The RF contains many trees, it randomly selects attributes as inputs. This algorithm creates a group of DT and a random subgroup of data using a bagging method. This classifier is considered as very efficient for finding solutions to any prediction problem. It can be used for problems related to regression and classification. It contains a combination of trees where eah tree gives output of a random sample with equal partition to all the trees inside the forest. The classifier will include two steps. First, ‘j’ random trees are generated, which makes a random forest. Second, it makes all the DTs which have the same features. Finally, the DTs which appear frequently or assessment of every DT are predicted.

## 4.3 Swarm intelligence

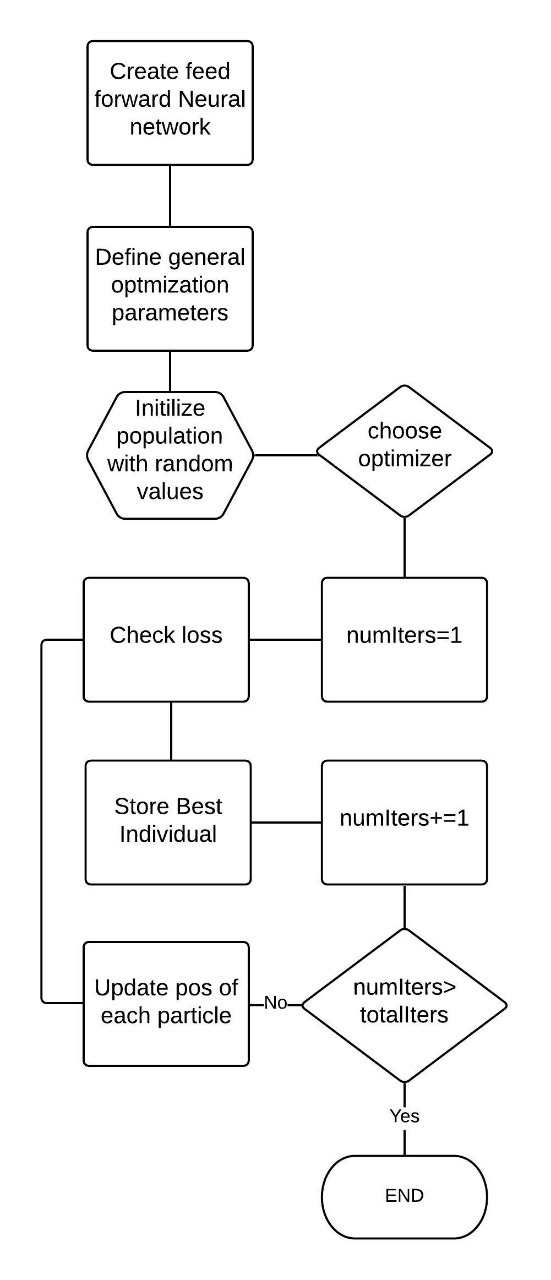
Swarm Intelligence(SI) deals with metaheuristic algorithms inspired by nature and population algorithms with the ability to solve complex problems by providing quick, stable and low-cost answers. Swarm intelligence is a subpart of artificial intelligence(AI) that includes clustered insect activities including birds, ants, honey bees and many others. These creepy crawlies share data straightforwardly or in a roundabout way among one

another utilizing a particular example for their endurance. The direct communication involves via video or audio, like honey bees communicate with flutter dance in a direct way. The indirect interaction is done when an insect makes a change to the system and the other insects detect the changed system, also communication of ants is an indirect way they leave pheromone chemical in their way to search food resources. Marco *et al*[17] explained swarm intelligence in a simple way. SI is generally used with optimization problems having large search space. We used some most popular algorithms in SI like

***Particle Swarm Optimization(PSO)*:** It solves the optimization problem by no. of iterations, in each iteration it tries to improve the global solution by using a population of individual solutions[18].

(1)

(2)



**Figure 3**

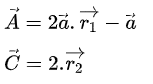
The equation 1 gives the velocity of the next particle and equation 2 determines pos. of the next particle where vkd(t) is the velocity of particle k in dim. d, t is for iterations. c1 and c2 are constants for tuning the exploration and exploitation. xkd(t) is the pos. of particle k in dimension d in iteration t. pkd(t) is the pos. of the personal best of particle k in dim d in iteration t. pgd(t) is the pos. of the global best of particle k in dim d in iteration t. R1, R2 are random numbers in between 0 and 1.

***Grey Wolf Optimization(GWO)*:** It imitates the grey wolf technique of hunting and governance. It employs four different gray wolf types(lowest to highest), Omega, Delta, Beta and Alpha . Also, it performs 3 steps in hunting the prey, searching, encircling and attacking[19].

(3)

(4)

Equation 3 is used to describe encircling behaviour of the wolves, Equation 4 to determine the next position of wolf. Where current iteration is denoted by t , coefficients ‘A’ and ‘C’ are vectors, their values can be calculated as shown in equation 5 where variable ‘a’ reduces from ‘2’ to ‘0’ linearly with iterations. ‘r1’ and ’r2’ are also vectors randomly between [0,1].

(5)

***Moth Flame Optimization(MFO)*:** In this algorithm individual solutions of the population are assumed as moths position. It can be in one-dimensional, two-dimensional or n-dimensional[20].

(6)

(7)

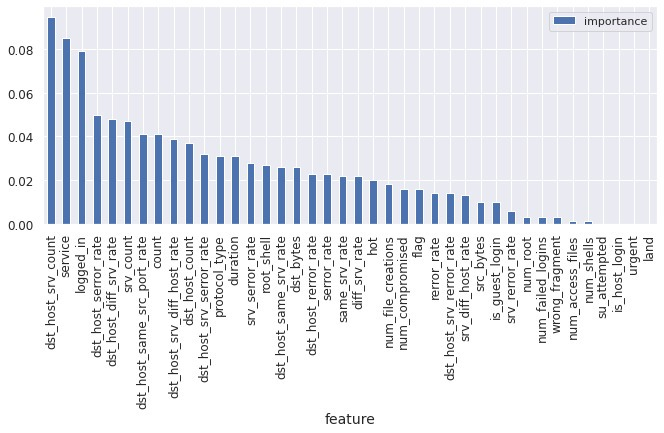
Equation 6 determines logarithmic spiral and Eq. 7 determines distance where ‘di’ determines the distance between moth ‘i’ and flame ‘j’, b is a constant that tunes the spiral shape, constant t lies in between -1 and 1. ‘Di’ determines the distance between ith moth and jth flame, Mi determines moth i, Fj determines the flame j.

***Multi Verse Optimizer(MVO)*:** It is mainly based on 3 cosmology concepts. They are black hole, white hole, wormhole, they are used to develop respective mathematical models of exploitation, exploration and local search.[20]

Figure3 outlines the methodology of Swarm Intelligence that has been applied to obtain the results for the Intrusion detection system.

# 5 Results

Nilesh et al[14] showed that 10 is the optimum number of features that can be used to predict an attack. Random Forest classifier has been applied to determine these ten features. The importance of the features can be seen in figure 4.



**Figure 4**

The selected features, as chose by RFC are,

'dst\_bytes','logged\_in', 'count','srv\_count','dst\_host\_srv\_count', 'dst\_host\_diff\_srv\_rate', 'dst\_host\_same\_src\_port\_rate', 'dst\_host\_serror\_rate', 'service', 'flag'.

Using the selected features, a feed forward neural network has been trained with Grey wolf optimization algorithm. The optimizer was chosen with the following parameters.

Number of search Agents: 1000

No of iterations: 25

lower bound= -10

upperbound=10

The performance of the algorithm is as follows.

Accuracy: 95.8 %

Detection rate: 97.5 %

false alarm rate: 2.8%

Precision: 0.9121929123089043

Recall: 0.8927171346687852,

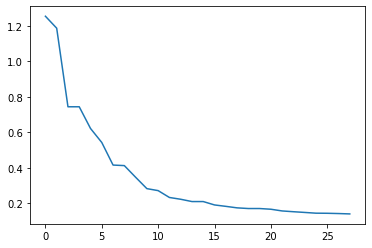
F-score: 0.9019993923408116

The loss on the training set is 0.1394.

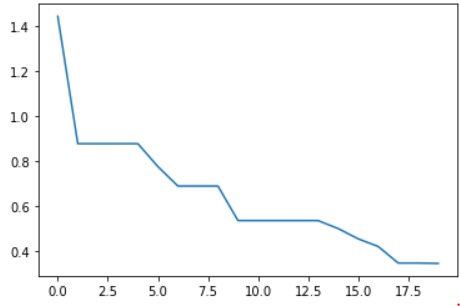
# 6 Performance evaluation

A comparative study was made to determine the best algorithm for IDS.

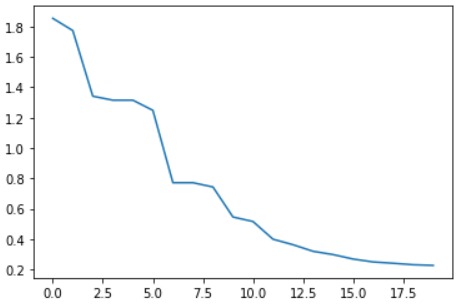
Figure 5, Figure 6, Figure 7 and Figure 8 shows the loss obtained on the training set on the Y-axis and iterations on the X-axis, for GWO, PSO, MVO and MFO respectively.

****

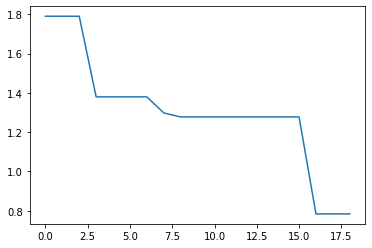
**Figure 5 : Grey wolf optimization**

****

**Figure 6: Particle Swarm optimization**

****

**Figure 7: Multi-verse optimization**

****

**Figure 8: Moth flame optimization**

Figure 9 shows the comparison between performance of all the above algorithms. The metric used is accuracy and detection rate.

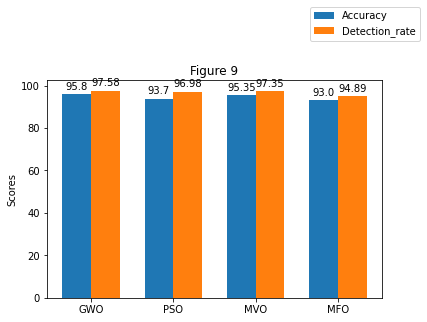
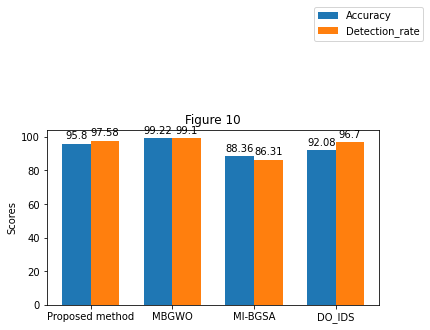


Figure 10 shows a comparative analysis of the proposed method with results achieved by other researchers, who have implemented IDS using machine learning and nature-inspired meta-heuristic algorithms: MBGWO[23], MI-BGSA [24] and DO\_IDS [25].



# 7 Conclusion and future work

This study describes the implementation of the PSO, GWO,MFO and MVO algorithm with feature selection in order to boost IDS performance and detection rate . We used only ten attributes from the NSL-KDD dataset in order to eliminate all irrelevant and least wanted ones that have harmful encounters on system achievement. Dataset simplification is the main aim of feature selection method by decreasing its level of complexity and identifying the useful set of features. The random forest classifier is applied for choosing the top 10 in the 41 attributes. By considering the complete study we conclude that GWO outperforms well.The accuracy and other measurements of success observed are higher than remaining algorithms.

More experimentation can be done with the proposed method by trying different neural network architectures. More meta-heuristic algorithms need to be tried and tested. Cross Validation can be performed on many more algorithms to choose the best algorithm which can improve the performance.

# 8 Acknowledgements

To achieve the success of this project, we got very good guidance and assistance. We took help from Nilesh[14] to choose Random Forest classifier for performing Feature Selection. Dr. Vijay helped us in making the research of all metaheuristic algorithms for minimizing the error. We would like to thank Dr. Kuldeep, who gave this opportunity for making this research which helped us to know new things.

# REFERENCES

[1] Aghdam M H and Kabiri P 2016 Feature selection for intrusion detection system using ant colony optimization. IJ Netw. Secur. 18(3): 420–432

[2] Myerson J M 2002 Identifying enterprise network vulnerabilities. Int. J. Netw. Manag. 12(3): 135–144

[3] Bhuyan M H, Bhattacharyya D K and Kalita J K 2014 Network anomaly detection: methods, systems and tools. IEEE Commun. Surv. Tutor. 16(1): 303–336

[4] Teodoro P G, Verdejo J D, Fernandez G M and Vazquez E 2009 Anomaly-based network intrusion detection:

techniques, systems and challenges. Comput. Secur. 28(2): 18–28

[5] Xue B, Zhang M and Browne W N 2013 Particle swarm optimization for feature selection in classification: a multiobjective approach. IEEE Trans. Cybern. 43(6): 1656–1671

[6] Xue B, Zhang M and Browne W N 2014 Particle swarm optimisation for feature selection in classification: novel initialisation and updating mechanisms. Appl. Soft Comput. 18: 261–276

[7]Tavallaee M, Bagheri E, Lu W and Ghorbani A A 2009 A detailed analysis of the KDD cup 99 data set. In: Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications, pp. 1–6

[8] Ganapathy S, Kulothungan K, Muthuraj Kumar S, Vijayalakshmi M, Yogesh P and Kannan A 2013 Intelligent feature selection and classification techniques for intrusion detection in networks: a survey. EURASIP J. Wirel. Commun. Netw. 1: 242–255

[9] Ahmad I and Amin F 2014 Towards feature subset selection in intrusion detection. In: Proceedings of the 7th IEEE Joint International Information Technology and Artificial Intelligence Conference, pp. 68–73

[10]Ian, Abraham A and Thomas J P 2005 Feature deduction and ensemble design of intrusion detection systems. Comput. Secur. 24(4): 295–307

[11]Tsai C F, Hsu Y F, Lin C Y and Lin W Y 2009 Intrusion detection by machine learning: a review. Expert Syst. Appl. 36(10): 11994–12000

[12]Bharathy A M V and Basha A M 2017 A multi-class classification MCLP model with particle swarm optimization for network intrusion detection. Sadhana 42(5): 631–640

[13]Sung A H and Mukkamala S 2003 Identifying important features for intrusion detection using support vector machines and neural networks. In: Proceedings of the Symposium on Applications and the Internet, pp. 209–216

[14]Kunhare, N., Tiwari, R. and Dhar, J., 2020. Particle swarm optimization and feature selection for intrusion detection system. *Sadhana*, *45*(1).

[15] Zhang H and Singer B H 2010 Recursive partitioning and applications. Springer Science & Business Media, pp. 1–258.

[16] Liaw A, Wiener M 2002 Classification and regression by random forest. R News 2(3): 18–22

[17] Bonabeau E, Marco D R D F, Dorigo M and Theraulaz G 1999 Swarm intelligence: from natural to artificial systems. Oxford University Press, vol. 1, pp. 1–320

[18]Kennedy, J. and Eberhart, R., 1995, November. Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks* (Vol. 4, pp. 1942-1948). IEEE.

[19]Mirjalili, S., Mirjalili, S.M. and Lewis, A., 2014. Grey wolf optimizer. *Advances in engineering software*, *69*, pp.46-61.

[20]Mirjalili, S., 2015. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-based systems*, *89*, pp.228-249.

[21]Franco E D L H, Garcia A O, Lopera J O, Correa E D L H and Palechor M F 2015 Implementation of an intrusion detection system based on self organizing map. J. Theor. Appl. Inf. Technol. 71(3): 324–334

[22]Kubat, M. and Matwin, S., 1997, July. Addressing the curse of imbalanced training sets: one-sided selection. In *Icml* (Vol. 97, pp. 179-186).

[23]Alzubi Q M, Anbar M, Alqattan Z N M, Al-Betar M A and Abdullah R 2019 Intrusion detection system based on a modified binary grey wolf optimisation. Neural Comput.Appl., pp. 1–13

[24] Bostani H and Sheikhan M 2017 Hybrid of binary gravitational search algorithm and mutual information for feature selection in intrusion detection systems. Soft. Comput. 21(9): 2307–2324

[25] Ren J, Guo J, Qian W, Yuan H, Hao X and Jingjing H 2019 Building an effective intrusion detection system by using hybrid data optimization based on machine learning algorithms. Secur. Commun. Netw.s, pp. 1–12