International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

Hybrid Classification Technique for Accurate Detection of DDoS Attacks

B. Satya Sai Vani.V, Shashi.M

Abstract*: Intellectual intrusion detection system can merely be build if there is accessibility to an effectual data set. A high dimensional quality dataset that imitates the real time traffic facilitates training and testing an intrusion detection system. Since it is complex to scrutinize and extort knowledge from high-dimensional data, it is identified that feature selection is a preprocessing phase during attack defense. It increases the classification accuracy and reduces computational complexity by extracting important features from original data. Optimization schemes can be utilized on the dataset for selecting the features to find the appropriate subspace of features while preserving ample accuracy rate characterized by the inventive feature set. This paper aims at implementing the hybrid algorithm, ABC-LVQ. The bio-inspired algorithm, Artificial Bee Colony (ABC) is adapted to lessen the amount of features to build a dataset on which a supervised classification algorithm, Linear Vector Quantization (LVQ) is applied, thus achieving highest classification accuracy compared to k-NN and LVQ. The NSL-KDD dataset is scrutinized to learn the efficiency of the proposed algorithm in identifying the abnormalities in traffic samples within a specific network.

Keywords* : *Artificial Bee Colony, Classification, DDoS Attacks, LVQ, Optimization.

I. INTRODUCTION

Now-a-days, DoS or DDoS attack detection has attracted many researchers worldwide. Attack detection techniques have been developed in order to protect network against misbehaving users. Such techniques have been continually improved in order to boost their detection capability. Data-centric approach such as knowledge discovery or data mining is one of popular method which has gained a lot of attention in many areas. Data mining is a practice of digging up out of sight prototype knowledge from outsized databases using mathematical, statistical, machine learning and AI techniques [3] [4].

Feature selection, an often used preprocessing technique for effectual data analysis in the promising field of data mining aims to choose a subset of unique features to optimally reduce the attribute space.

Revised Manuscript Received on February 28, 2020.
***** Correspondence Author
B. Satya Sai Vani. V*, Research Scholar, Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam (Andhra Pradesh) India.
Dr. Shashi.M, Professor, Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam (Andhra Pradesh) India.

In current years, feature selection has been lucratively applied in classification problem, like information retrieval processing, data mining applications and pattern classification.

One of an efficient and robust feature selection approaches is Artificial Bee Colony (ABC) algorithm that is stimulated by surveillance on real ants in their search for the shortest paths to food sources.

The author compose the use of ABC technique to effectively select optimized feature subset that are more relevant to the Linear Vector Quantization classification task to obtain higher classification accuracy compared to accessible algorithms. NSL-KDD dataset was employed to estimate the efficacy of this proposed algorithm. Experimental results demonstrate that, this adapted approach can achieve significant detection accuracy thus reducing the misclassification instances, the error rate, and computational complexity. The manuscript is structured below in four sections. Section II confers some related work. Section III presents a concise depiction of NSS-KDD dataset and ground rules about the ABC algorithm and LVQ algorithms. It presents the proposed ABC-LVQ classification method in detail. Investigational outcomes are offered in Section IV to exhibit the ability of the proposed ABC-LVQ on classification of intrusions using NSL-KDD dataset. Conclusion part and Future scope are drawn in Section-V.

II. LITERATURE REVIEW

To select features, few intelligent optimization algorithms were proposed such like genetic algorithm (GA). particle swarm optimization (PSO), ant colony optimization (ACO) and artificial bee colony (ABC) [1-4]. These techniques help to choose a little informative or tangible feature variable to improve accurate and efficient data analysis. ABC algorithm gained a large amount attention owing to its Control parameters and greater optimization performance [5][6].

In [7], a hybrid approach based on artificial neural networks (ANN) and ABC algorithm was offered to opt for feature subset successfully. Schiezaro and Pedrini [4] anticipated a feature assortment method with ABC algorithm for categorizing different UCI datasets where the selected feature set could offer better classification accuracy. Selecting features [8] utilizing ABC and Nearest Neighbor were introduced for steganalysis in images. Support Vector Machine (SVM) stood prevailing in sorting compared to -NN classifier and ANN by reason of its exceptional classification accuracy and generalization act [9][10].

Alternatives of ABC algorithm [11] were projected to progress the overall seek performance and convergence rate as well, that focuses mainly on the initialization of population and strategy of solution search [12].

GABC, the gbest-guide ABC algorithm was introduced by Zhu and Kwong [13] that integrates information of global preeminent solution and solution search equation to develop the exploitation. Researchers verified that the HGABC algorithm has the prevailing capability of probing universal optimal solution to standard optimization algorithm, GABC. [20].

Wei-feng et al., [14] proposed a tailored ABC (MABC) algorithm by initiating the chaotic and opposition-based initialization along with the finest solution of the preceding iteration.

To advance the utilization and keep the investigation of ABC, Zhang and Liu [15] projected a new ABC (NABC) algorithm together with the global preeminent solution and a arbitrary solution into the search equations of the onlookers and employed bees, correspondingly.

Zhen-an et al. [16] initiated a SDABC algorithm from three facets that includes initialization of search space division, disruptive selection strategy, and enhanced scout bee phase.

Advanced edition of the KDD cup99 dataset is NSL-KDD dataset [5]. Different types of analyses were performed by Researchers on NSL-KDD dataset by utilizing variety of tools and techniques with a collective objective of developing an effectual intrusion detection system. A thorough scrutiny on NSL-KDD dataset with an array of machine learning techniques was made [6] by means of WEKA tool.

K-means clustering algorithm has been applied on NSL-KDD dataset [7] to train and test different open and new attacks. Self Organization Map (SOM) Artificial Neural Network was employed to compare NSL-KDD dataset with its ancestor KDD99 cup dataset [8]. KDD99, GureKDD and NSLKDD datasets were utilized to perform exhaustive analysis on use of data mining supported machine learning techniques like Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), K-Means and Fuzzy C-Mean clustering algorithms.

Prasad et al. [8] had classified DDoS detection methods as statistical, knowledge based, soft computing, and data mining & machine learning. On the other hand, the authors categorized different methods depending on centralized and distributed modes. Different ways to categorize detection methods were cited by Douligeris and Mitrokotsa and Peng in terms of DoS-specific and anomaly detection [10] [11].

Karim et al. [12] stated that, there is a great inclination in use of data mining and machine learning methodologies For example, (SVM) support vector machine, neural network (NN), and cluster analysis. Moreover, there is also an improved concern on detecting DDoS detection by a distributed mode [3,13]. In précis, the recent tendency of DDoS detection takes into account an increase of traffic and the insight that can be produced by the availability of data.

Simulation was adopted to contrast the accuracy and efficiency of diverse models in perceiving DDoS. Most of the researchers employed simulation to compute the accuracy of dissimilar learning models in detecting DDoS attacks. Percentage of true positive (TN) , true negative (TN), false positive (FP) and false negative (FN) were calculated for this purpose. It was shown that computing time for both training and testing can be a good indicator to evaluate the efficiency and the performance of a model. [3,6,9,14-15].

III. RESEARCH METHODOLOGY

This The dataset, NSL-KDD was utilized to divulge the most defenseless protocol that is often used by intruders to instigate intrusions within a network. The dataset consists of 125974 records with 41 different attributes. ABC algorithm is applied on KDD dataset to get the optimized dataset that is trained to LVQ for structuring a model. Again 40% of the KDD dataset is taken as test dataset and applied on the model to classify the accurate DDoS attack.

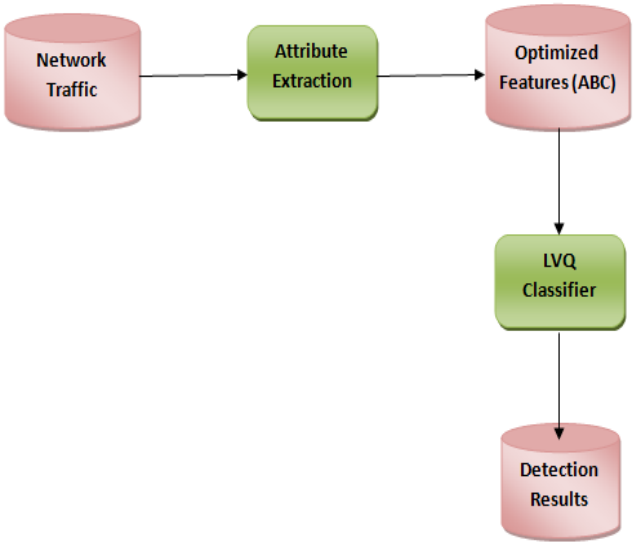

Figure 1: Proposed Workflow

A. Feature Selection

Each observation has 41 attributes unfolding diverse features of the surge and a class label indicating either normal or attack. To determine the most relevant subset of features from the data, an optimization method, Artificial Bee Colony is used while preserving adequate precision rate of inventive features.

Algorithm 1: Artificial Bee Colony Algorithm

Initialize all Parameters
Repeat Step 1 to Step 3 till the termination criteria meet
Step 1: Apply employed bee phase to generate new food sources using(2)
Step 2: Onlooker bee phase to update food sources according to their nectar quality
Step 3: Scout bee phase to generate new solution in place of rejected solutions
Remember the finest solutions established until now

Figure 2: Artificial Bee Colony Algorithm

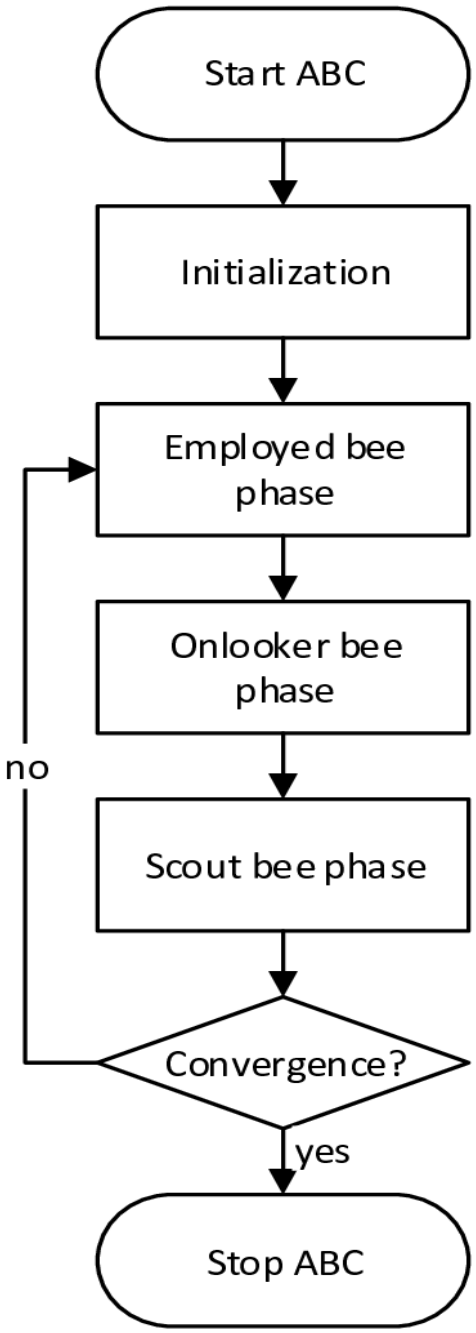

Figure 3: Flow of ABC Algorithm

An algorithm, Artificial Bee Colony is an intellectual method stirred by foraging performance of a bee colony and extensively utilized to unravel incessant numerical optimization difficulties [17]. In ABC method, the bee colony comprises of three varieties of bees: scout, employment, also onlooker bees. Through the progression of optimization, the site of a food originator implies a probable solution for an optimization difficulty. ABC is an iterative practice that carries out recurring searches for the solution with all the three categories of bees until the utmost number of cycles reaches, MAXcycles [18, 19].

B. Linear Vector Quantization

Learning vector quantization (LVQ) is one of the algorithms that are widely used in classifying potentially high dimensional data. LVQ procedures are effortless to execute and instinctively unambiguous. The LVQ classification depends upon a distance calculation, typically a Euclidean distance that quantifies the resemblance of particular data with supposed prototype or codebook vectors, those represents classes. The models are resolute in a training process through labeled instance data and are construed in a undemanding way as they capture important features in the data in same space. LVQ classification algorithm supports binary as well as multi-class classification problems. The intricacy of LVQ network can be restricted through training according to the definite requirements.

Algorithm

1. LVQ is a group of codebook vectors.
2. A codebook vector comprises of a set of numbers having I/O attributes same as the training data. As an instance, if the problem contains length, height and width as attributes, then even a codebook vector also comprises of the same attributes.
3. A set of codebook vectors erudite from training_data are clubbed to form the model. They seem to be training illustrations, however the value of all attributes have been tailored depending on the learning practices.

IV. EXPERIMENTAL RESULTS

The Proposed system has been implemented using R Studio. The dataset consists of 125974 records with 41 different attributes. The dataset necessitate preprocessing primarily to optimize thus reducing the irrelevant features and noisy data. Classification is done in this work by using LVQ algorithm. Analysis is done to measure the efficacy of classification algorithms in grading the NSL-KDD dataset. Rate of accuracy in perceiving the normal and attack class of system in terms of time and space complexities are shown below.

You can use color figures as per the requirement but fonts should be in black. Authors can use any number of color diagram, chart, picture, screenshots, and any snap which is required for the research of the title.

src_bytes	dst_bytes	logged_in	count	srv_count	serror_rate	srv_serror_rate	rerror_rate	srv_rerror_rate	same_srv_rate	diff_srv_rate	srv_diff_host_rate	dst_host_count	dst_host_srv_count	dst_host_same_srv_rate	dst_host_diff_srv_rate	dst_host_same_src_port_rate	dst_host_srv_diff_host_rate	dst_host_serror_rate	dst_host_srv_serror_rate	dst_host_rerror_rate	dst_host_srv_rerror_rate	Attack Type
491	0	0	2	2	0	0	0	0	1	0	0	150	25	0.17	0.03	0.17	0	0	0	0.05	0	normal
146	0	0	13	1	0	0	0	0	0.08	0.15	0	255	1	0	0.6	0.88	0	0	0	0	0	normal
0	0	0	123	6	1	1	0	0	0.05	0.07	0	255	26	0.1	0.05	0	0	1	1	0	0	neptune
232	8153	1	5	5	0.2	0.2	0	0	1	0	0	30	255	1	0	0.03	0.04	0.03	0.01	0	0.01	normal
199	420	1	30	32	0	0	0	0	1	0	0.09	255	255	1	0	0	0	0	0	0	0	normal
0	0	0	121	19	0	0	1	1	0.16	0.06	0	255	19	0.07	0.07	0	0	0	0	1	1	neptune
0	0	0	166	9	1	1	0	0	0.05	0.06	0	255	9	0.04	0.05	0	0	1	1	0	0	neptune
0	0	0	117	16	1	1	0	0	0.14	0.06	0	255	15	0.06	0.07	0	0	1	1	0	0	neptune
0	0	0	270	23	1	1	0	0	0.09	0.05	0	255	23	0.09	0.05	0	0	1	1	0	0	neptune
0	0	0	133	8	1	1	0	0	0.06	0.06	0	255	13	0.05	0.06	0	0	1	1	0	0	neptune
0	0	0	205	12	0	0	1	1	0.06	0.06	0	255	12	0.05	0.07	0	0	0	0	1	1	neptune
0	0	0	199	3	1	1	0	0	0.02	0.06	0	255	13	0.05	0.07	0	0	1	1	0	0	neptune
287	2251	1	3	7	0	0	0	0	1	0	0.43	8	219	1	0	0.12	0.03	0	0	0	0	normal
334	0	1	2	2	0	0	0	0	1	0	0	2	20	1	0	1	0.2	0	0	0	0	warezclient
0	0	0	233	1	1	1	0	0	0	0.06	0	255	1	0	0.07	0	0	1	1	0	0	neptune
0	0	0	96	16	1	1	0	0	0.17	0.05	0	255	2	0.01	0.06	0	0	1	1	0	0	neptune
300	13788	1	8	9	0	0.11	0	0	1	0	0.22	91	255	1	0	0.01	0.02	0	0	0	0	normal
18	0	0	1	1	0	0	0	0	1	0	0	1	16	1	0	1	1	0	0	0	0	ipsweep

Figure 4: Optimized NSL-KDD dataset after application of ABC algorithm

SNO	Algorithms	Time Complexity(msec)
1	Optimized LVQ	12456
2	LVQ	25688
3	kNN	35688

Table 1: Shows the comparison of time complexity of Optimized LVQ with LVQ and kNN

SNO	Algorithms	Data Records	Space Complexity(kb)
1	Optimized LVQ	125974	12254
2	LVQ	125974	36988
3	kNN	125974	69988

Table 2: Shows the comparison of space complexity of Optimized LVQ with LVQ and kNN

SNO	Algorithms	Error Rate
1	Optimized LVQ	0.3289
2	LVQ	0.589
3	kNN	0.788

Table 3: Shows the error rate of Optimized LVQ with LVQ and kNN

Attacks	ipsweep	neptune	nmap	normal	portsweep	satan	smurf
Frequency	3599	41214	1493	67343	2931	3633	2646

Table 4: Shows the frequency of Attacks by type

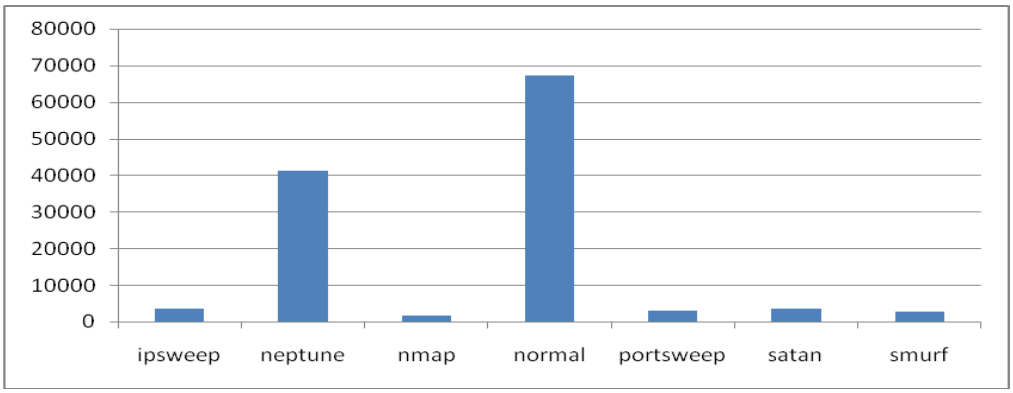

Figure 5: Graph showing the Frequent Attacks

V. CONCLUSION

The scope of this study is to find the potential features of DDoS attacks to explore its classification accuracy. To accomplish this objective, a model is anticipated where ABC and LVQ were applied consecutively on NSL KDD dataset. The proposed method provides a way to optimize the possible attributes for the DDoS attack detection. From the experimental results it is proved that optimized ABC-LVQ is most appropriate for classifying the DDoS attacks. It gives accurate 96% true classification rate which is relatively higher than kNN and LVQ. To bring to a close, it can be assumed that optimum classification can be obtained by preferring the optimized LVQ to other existing algorithms. Moreover, this revision is an effort to differentiate the performance of the algorithms with respect to time and space complexities and accuracy rate to give the accurate prediction of attack type for signifying appropriate actions accordingly.

REFERENCES

1. Zhao, Mingyuan, et al. "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes." Expert Systems with Applications 38.5 (2011): 5197-5204.
2. Xue, Bing, Mengjie Zhang, and Will N. Browne. "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms." Applied soft computing 18 (2014): 261-276.
3. Chen, Bolun, Ling Chen, and Yixin Chen. "Efficient ant colony
4. optimization for image feature selection." Signal processing 93.6 (2013): 1566-1576.
5. Schiezaro, Mauricio, and Helio Pedrini. "Data feature selection based on Artificial Bee Colony algorithm." EURASIP Journal on Image and Video processing 2013.1 (2013): 47.
6. Akay, Bahriye, and Dervis Karaboga. "A survey on the applications of artificial bee colony in signal, image, and video processing." Signal, Image and Video Processing 9.4 (2015): 967-990.
7. Jia, Dongli, Xintao Duan, and Muhammad Khurram Khan. "Modified artificial bee colony optimization with block perturbation strategy." Engineering Optimization 47.5 (2015): 642-655.
8. Shokouhifar, Mohammad, and Shima Sabet. "A hybrid approach for effective feature selection using neural networks and artificial bee colony optimization." 3rd international conference on machine vision (ICMV 2010).
9. Mohammadi, Farid Ghareh, and Mohammad Saniee Abadeh. "A new metaheuristic feature subset selection approach for image steganalysis." Journal of Intelligent & Fuzzy Systems 27.3 (2014): 1445-1455.
10. Zhang, Xueying, Xiaofeng Liu, and Zizhong John Wang. "Evaluation of a set of new ORF kernel functions of SVM for speech recognition." Engineering Applications of Artificial Intelligence 26.10 (2013): 2574-2580.
11. Moosavian, Ashkan, Hojat Ahmadi, and Ahmad Tabatabaeefar. "814. Fault diagnosis of main engine journal bearing based on vibration analysis using Fisher linear discriminant, K-nearest neighbor and support vector machine." Journal of Vibroengineering 14.2 (2012).
12. Karaboga, Dervis, et al. "A comprehensive survey: artificial bee colony (ABC) algorithm and applications." Artificial Intelligence Review 42.1 (2014): 21-57.
13. Duan, Haibin, and Qinan Luo. "New progresses in swarm intelligence-based computation." International Journal of Bio-Inspired Computation 7.1 (2015): 26-35.
14. Zhu, Guopu, and Sam Kwong. "Gbest-guided artificial bee colony algorithm for numerical function optimization." Applied mathematics and computation 217.7 (2010): 3166-3173.
15. Gao, Wei-feng, and San-yang Liu. "A modified artificial bee colony algorithm." Computers & Operations Research 39.3 (2012): 687-697.
16. Zhang, Song, and Sanyang Liu. "A novel artificial bee colony algorithm for function optimization." Mathematical Problems in Engineering 2015 (2015).
17. He, Zhen-an, et al. "A modified artificial bee colony algorithm based on search space division and disruptive selection strategy." Mathematical Problems in Engineering 2014 (2014).
18. Karaboga, Dervis. An idea based on honey bee swarm for numerical optimization. Vol. 200. Technical report-tr06, Erciyes university, engineering faculty, computer engineering department, 2005.
19. Karaboga, Dervis, and Bahriye Basturk. "On the performance of artificial bee colony (ABC) algorithm." Applied soft computing 8.1 (2008): 687-697.
20. Akay, Bahriye, and Dervis Karaboga. "A modified artificial bee colony algorithm for real-parameter optimization." Information sciences 192 (2012): 120-142.
21. Shah, Habib, et al. "Hybrid guided artificial bee colony algorithm for numerical function optimization." International Conference in Swarm Intelligence. Springer, Cham, 2014.

AUTHORS PROFILE

Satyasaivani. B, is an Associate Professor in Department of Computer Science at GITAM (Deemed to be University) , Visakhapatnam, India. Her research interests include Cyber Security, Machine Learning, Information Retrieval and Data Mining. She has more than 20 years of teaching and research experience. She has published 8 publications in International and National journals. She also served as a member in several seminars, conferences and workshops. She is a life time member of Indian Science Congress Association.

Dr. Shashi. M, is a Professor in the Department of Computer Science & Systems Engineering, A.U. College of Engineering(A), Andhra University, Visakhapatnam, Andhra Pradesh since 1999. She received the AICTE Career Award in 1996, Best Ph.D thesis prize from Andhra University in the year 1994 and AP State Best teacher award in 2016. 13 Ph.D.'s were awarded under her guidance. She co-authored more than 60 technical research papers in International Journals and 50 International Conferences and delivered many invited talks in such academic events. She is a member of IEEE Computational Intelligence group, Fellow of Institute of Engineers (India) and life member of Computer Society of India.. Her current research interests include Data Mining, Data Analytics, Artificial Intelligence, Pattern Recognition, Soft Computing and Machine Learning.