Detecting Intrusion Behavior in Communication Networks using Firefly-based Fuzzy Clustering Approach

D. Veeraiah, Tejaswi Kavuru, Ebenezer Jangam, P. Victer Paul

ABSTRACT--- Intrusion detection system is responsible to identify any suspicious activity in a communication network. Researchers proposed diverse methods for intrusion detection as it is a necessary task to provide security for the communication network and users. In this paper, intru-sion detection system is proposed using the combination of Firefly algorithm and fuzzy clustering. Initially, firefly algorithm is used to optimize the separation between the clusters. The output of firefly algorithm is given to the fuzzy clustering. Fuzzy clustering is used to differentiate the malicious activity from the normal activity. The proposed approach is evaluated on benchmark IDS datasets and the results are encouraging.

Keywords: Fuzzy C-Means (FCM), Firefly Algorithm (FA) and Intrusion Detection System (IDS).

I. INTRODUCTION

Dorothy E. Dinning in 1987 proposed a proposal to identify intruders as a method of responding computer and networking attacks and abuse. Intrusion detection is applied by an intrusion detection system, and today there are many profitable intrusion detection systems are existing. In overall, most of these profitable applications are relatively active and are unsatisfactory, requiring extra research into the systems that find additional dynamic intruders. Files and records are the resources for which attackers look to access without authorization. It's hard to safeguard all the records and files that are simply accessible by various attackers via the web. One way is to find out the attacker or intruder as soon as malicious activity is recorded. The Intrusion Detection System (IDS) [2]-[4] was developed in order to detect malicious activity and to trace out the intruder with malicious intention. IDS issues warning when malicious activity is detected in the network. IDS can be designed in different ways. In general, the IDS follow the two-step process. The first consists of various tasks such as a system configuration inspection to find host-based and offensive settings; Finding confusing pass-words without checking password words; and check several other system parts to identify policy violations. The second is a network-based. Nowadays, computers are connected to networks that are quickly accessible and the network intrusions quickly.

One of the ways to design IDS is by using clustering approach. Clustering approaches can be used to identify abnormal activities. Once an activity is identified as malicious activity, it is reported immediately so that suitable action can be taken against the intruder. Clustering is to organize a cluster of objects in to two different groups according to their similarities. One group contains objects with normal behavior and other group contains objects with abnormal behavior. One of the most common clustering techniques is fuzzy c-means clustering. Fuzzy c-means [1] clusters the data points such that a data point can belong to more than one cluster with a determined degree. This determined degree is called the membership value of data points and membership value remains in range (0, 1). Even though fuzzy c-means is one of the most commonly used techniques, there are some problems. The main problems associated with fuzzy c-means are (i) randomly generated cluster centers are used for initialization and (ii) probability of getting trapped in local minima is high.

In spite of the popularity of fuzzy c-means, the drawback of the algorithm is that it can be trapped at the local optimum rather than global optimum [4]. Firefly optimization was proposed by X S Yang. Firefly Algorithm [5] has a property of achieving a high rate of converging at global optimum.

IDS can be designed in different ways and researchers have proposed a variety of methods for intrusion detection. The figure 1 depicts different kinds of intrusion detection systems.

Many researchers have proposed different approaches, which are very effective in detecting unknown intrusions in a system. Design of Intrusion detection system varies from each other and it depends on factors like type of intruder, deployment method, data collection method and expected behavior of intruder. We propose IDS based on FCM optimized using Firefly algorithm. The proposed algorithm has been identified as potential and effective to identify intrusions.

This paper is organized as follows: Next section provides details about Fuzzy Clustering and Firefly Algorithm along with proposed approach Firefly-based Fuzzy Clustering. Section 3 explains Results and discussions. Section 4 concludes the paper.

Revised Manuscript Received on February 11 , 2019.

D. Veeraiah Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research Deemed to be University, Andhra Pradesh, India. (E-mail: d.veeraiah@gmail.com)

Tejaswi Kavuru Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research Deemed to be University, Andhra Pradesh, India. (E-mail: tejaswikavuru25@gmail.com)

Ebenezer Jangam Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research Deemed to be University, Andhra Pradesh, India. (E-mail: ebenezer.jangam@gmail.com)

P. Victer Paul Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research Deemed to be University, Andhra Pradesh, India. (E-mail: victerpaul@gmail.com)

II. PROPOSED WORK

2.1 Fuzzy Clustering

J.C. in 1973 Dunn's Fuzzy C-means (FCM) clustering was developed and in 1981 J.C. Developed by Bezdek. Intrusion detection system (FCM) algorithm [6] is used to identify the difference between a network's intrusions and normal activities. Several clusters can be distinguished by a FCM through a one piece of the data. This is a partition algorithm that w.r.t minimizes target function. Partition matrix (Eq. 1) [7].

$$J(U,V) = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} \left\|a_j - b_i\right\|^2 \quad (1)$$

$$U_{ij} = \left[\left(\sum_{k-1}^{c} \frac{\|a_j - b_i\|}{\|a_j - b_k\|}\right)^{\frac{2}{m-1}}\right]^{-1} \quad (2)$$

In Eq. 1, 'a_j' and 'b_i' are respectively cluster point and ith cluster centre. $U_{i,j}$ Represents the relationship value of the j_{th} cluster point w.r.t. Cluster 'i'. 'm' Fuzzy Controlling Factor. Its value is '1' and '∞.' Hard partition can be created to set the fuzzy and to set the value. Norm function is ‖ ‖. The 3 main factors used by FCM is mainly partition matrix, objective function and fuzzy membership function Eq.3.$u^m_{ij}$is intended as in eq.2 and cluster centre in Eq.3In eq. 3 , $i \geq 1$, $i \leq c$.

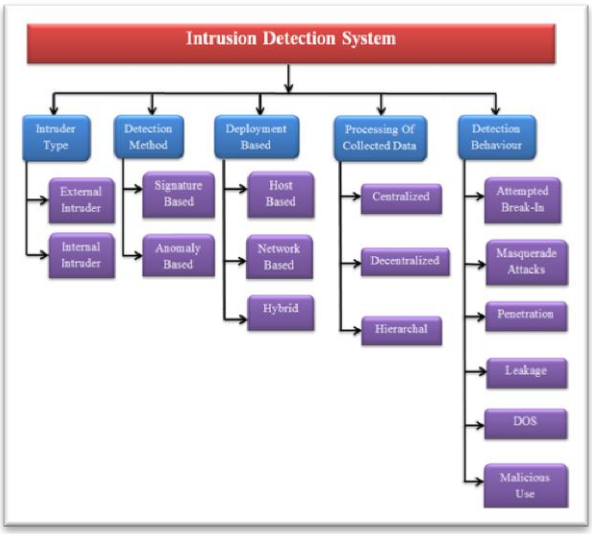

Fig: 1: Classification of IDS

$$V_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} a_j}{\sum_{j=1}^{N} u_{ij}^{m}} \quad (3)$$

The modification among K-means and fuzzy clustering is the u^m_{ij} (fuzzy factor). This fuzzy factor is responsible for the thickest level for different clusters.

Firefly algorithm is used to initialize the cluster centres to optimal values so that FCM can start with better cluster centres. FCM is used to differentiate normal and malicious activities. Firefly algorithm helps FCM to obtain optimal clusters. The selection of optimal clusters results in higher accuracy of the IDS. As Firefly algorithm iterates specified number of times, the computational overhead is feasible.

$$P = \{W_1, W_2 \ldots\ldots W_n\} \quad (4)$$

$$Wi = \{C_1^i, C_2^i \ldots\ldots C_m^i\} \quad (5)$$

$$C_j = \{V_1^j, V_2^j \ldots\ldots V_d^j\} \quad (6)$$

2.2.FIREFLY OVERVIEW

Firefly algorithm was created on the performance of fireflies and their movement. Three central concepts in Firefly algorithm are

- Any firefly can be attracted to any additional firefly irrespective of sex
- A firefly will be attracted towards a brighter firefly and less bright one moves to-wards brighter one
- The impartial purpose decides the intensity of the firefly

In simplified form, I(r) light intensity varies allowing to the inverse square law:

$$\mathbf{I(r)} = \frac{Is}{r2} \quad \mathbf{(7)}$$

The light intensity "I" and the interest coefficient "γ" varies with the space "r" for an average and it will be:

$$I = I_0 e^{-\gamma r} \quad (8)$$

I	= light intensity
I_0	= light intensity at first
γ	= light absorption constant
r	= space among firefly i and j

Where, i is the intensity of the unique light.

The mutual result of inverse squared law and immersion is regarded as a Gaussian form and is represented by:

$$\mathbf{I = I_0 e^{-\gamma r^2}} \quad \mathbf{(9)}$$

The attraction of Firefly is relative to the intensity of the light on the adjoining skins. The attraction of a firefly can be clear as "β":

$$\mathbf{B(r) = \beta_0 \gamma r 2} \quad \mathbf{(10)}$$

Where, "β_0" is attractiveness at r = 0 .

The Euclidean distance among the firefly "I" at x_i and firefly "j" at x_j is given by:

$$\mathbf{r_{ij}} = \sqrt{\sum_{k=1}^{d}(xik - xjk)\,2} \quad \mathbf{(11)}$$

Firefly "Eye" motion counting is another attractive (bright) firefly "j" and it estimated:

$$\mathbf{x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2}(x_j - x_i) + \alpha(rand - \frac{1}{2})} \quad \mathbf{(12)}$$

The second term refers to the attraction and the third word randomization. Anywhere "α" randomization parameter and "rand" is a random number creator, which is consistently spread among 0 and 1.

Algorithm -1: Proposed FA-FCM based Intrusion Detection

1. Initialize the population of n fireflies. Each firefly is initiated using random pair of cluster centres. One cluster is supposed to contain normal activity and another cluster is supposed to contain malicious activity.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

2. Then global best gbest is initialized with a firefly which is randomly chosen.
3. Intensity or attractiveness of each firefly is calculated based on the Eq.1
4. Compare the attractiveness for each pair of fireflies.
5. If attractiveness of i^{th} firefly is greater than j^{th} firefly in a given pair, then j^{th}firefly moves towards i^{th} firefly. Position of firefly is updated using Eq.12.
6. Update the attractiveness of fireflies.
7. Update gbest by comparing attractiveness of the fireflies from 1 to n
8. Repeat the steps from 3 to 7 until the stopping condition is met.
9. Initialize the FCM cluster centres using the value in gbest
10. Repeat FCM until convergence is achieved.
11. Update the membership function using Eq.1 at each step.
12. Calculate the cluster centres based on Eq.2 at each step.
13. If convergence is reached stop and finalize the clusters.

III. Results and Discussion

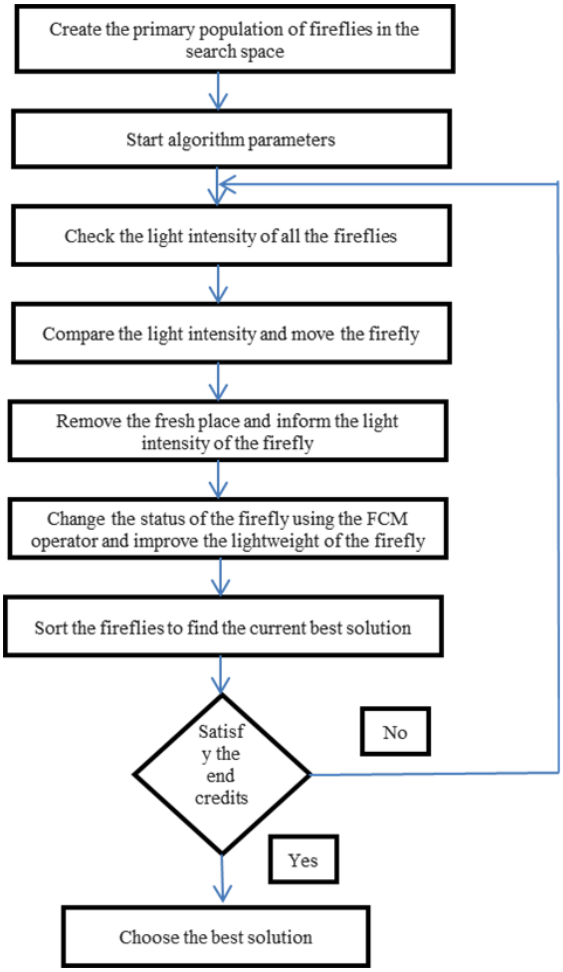

Fig. 2: Firefly-based Fuzzy Clustering approach

KDD Cup 99 dataset is used to evaluate the proposed approach. In edict to show the efficiency of the future approach, the performances of other clustering techniques are considered for intrusion detection.

The six clustering methods compared are Density-based Cluster, K-means, Filter Cluster, FCM, GA-FCM and FA-FCM. These study models showcases all activities in fig 3, the number of unrecognized activities and their percentages. The future FA-FCM IDS system was start to be better related to other models.

III. CONCLUSION

A hybrid approach using firefly algorithm and fuzzy clustering is proposed for intrusion detection in communication networks. The normal activities and malicious activities are separated from each other using the FA-FCM classifier. Using the standard IDS dataset, the performance of proposed approach is linked with five additional clustering techniques. The combination of FA and FCM has achieved better results when compared to FCM alone. Even though extra computation is involved in the proposed approach, it yields better results than FCM and FCM-GA..

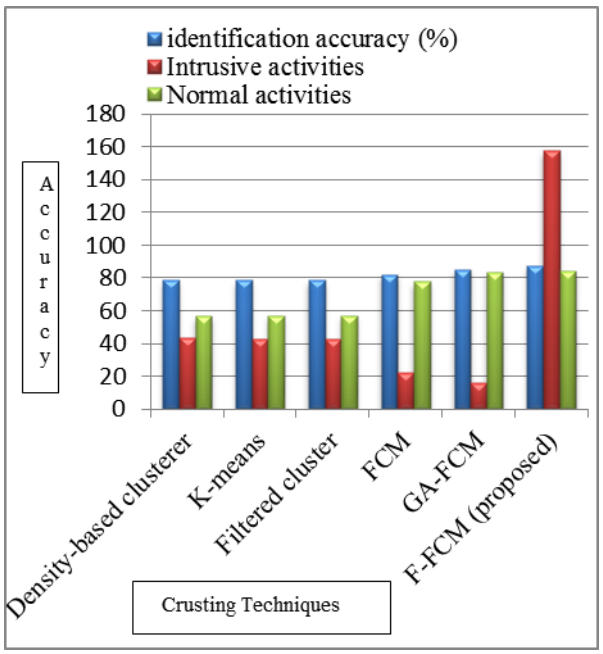

Fig. 3: Comparison Activities.

REFERENCES

1. Cai, W., Chen, S., Zhang, D," Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation" Pattern recognition, Volume 40, Issue 3, March 2007, Pages 825-838.
2. Mishra, Debasmita, and BighnarajNaik. "Detecting Intrusive Behaviors using Swarm-based Fuzzy Clustering Approach." Soft Computing in Data Analytics. Springer, 2018, pp 837-846.
3. Nayak, Janmenjoy, et al. "An improved firefly fuzzy c-means (FAFCM) algorithm for clustering real world data sets." Advanced Computing, Networking and Informatics-Volume 1. Springer, Cham, 2014. PP 339-348.
4. Zuech, R., Khoshgoftaar, T. M., & Wald, R., "Intrusion detection and big heterogeneous data: a survey", Journal of Big Data, 2015.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

5. Kumar, R., & Sharma, D, " HyINT: Signature-Anomaly Intrusion Detection System". In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE, 2018.
6. Özgür, A., &Erdem, H."A review of KDD99 dataset usage in intrusion detection and machine learning between 2010 and 2015". PeerJPrePrints, 2016.
7. Radhakrishna, V., Kumar, P. V., &Janaki, V. "Novel Similar Temporal System Call Pattern Mining for Efficient Intrusion Detection". J. Journal of Universal Computer Science 22(4):475-493 · July 2016.
8. Aissa, N. B., &Guerroumi, M. "A genetic clustering technique for Anomaly-based Intrusion Detection Systems". In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2015.
9. Wei Jiang ; Min Yao ; Jun Yan, "Intrusion Detection Based on Improved Fuzzy C-means Algorithm". In 2008 International Symposium on Information Science and Engieering, pp 326-329, IEEE,2008.
10. Asyali, M.H., Colak, D., Demirkaya, O., Inan, M.S. " Gene expression profile classification: a review" . Current Bioinformatics,pp 55-73, 2006.
11. Fister, I., Fister Jr., I., Yang, X.S. and Brest, J. " A Comprehensive Review of Firefly Algorithms". Swarm and Evolutionary Computation,Volume 13, December 2013, Pages 34-46.