



A New Concept of Network Intrusion Detection using Fuzzy Clustering

¹Richa Sampat, ²Shilpa Sonawani

^{1,2}Maharashtra Institute of Technology University of Pune, Maharashtra (India)
Email: ¹richasampat@gmail.com, ²shilpa.sonawani@mitpune.edu.in

Abstract—Internet has become a vital part of any organization. Sensitive and confidential information is being sent over the network. But with the growth of internet, intrusion and attacks have also increased. Thus, there arises a need of robust and powerful intrusion detection systems which can detect the attacks. Recently, many novel methods are experimented to build strong IDSs. The aim of this paper is to present a methodology that can recognise and detect attacks efficiently. In this paper, we implement the FCM algorithm and successfully integrate it in WEKA to expand the system functions of the open-source platform, so that users can directly call the FCM algorithm to do fuzzy clustering analysis. Besides, considering the shortcoming of the classical FCM algorithm in selecting the initial cluster centers, we represent an improved FCM algorithm which adopts a new strategy to optimize the selection of original cluster centers. A novel classification via dynamic fuzzy c means clustering algorithm has been proposed to build an efficient anomaly based network intrusion detection model. A subset of KDDCup 1999 intrusion detection benchmark dataset has been used for the experiment. The proposed novel concept will be efficient in terms of detection accuracy, low false positive rate in comparison to the other existing methods.

Keywords: Network Intrusion Detection, Fuzzy Clustering, Weka, Dynamic clustering, Dynamic Fuzzy C Means.

I. INTRODUCTION

With the growth of network technology, nowadays more and more people are learning new ways to attacks through the network resources and carry out extremely destructive attacks. In recent times, the amount of hackers attacks is growing at 10 times per year. Therefore, network security becomes a vital factor of computer technology. The concept of Intrusion detection system (IDS) proposed by Denning (1987) is useful to detect, identify and track the intruders [1]. Intrusion Detection is the process of monitoring and analyzing the events occurring in a computer system in order to detect signs of security problems.

Intrusion Detection is an important aspect as very large amount of sensitive information is stored and processed in networked systems across the globe. IDS modeling is a critical task and in recent years, various techniques of

data mining and soft computing are being explored to strengthen network security. These approaches include neural networks, decision trees, support vector machines, classification and clustering techniques, fuzzy logic, etc. These methods can detect both known and unknown patterns of attacks, thereby helping in development of smart intrusion detection systems. In this paper, the focus is on variation of fuzzy clustering technique that is used for building an intrusion detection system.

The rest of the paper is organized as follows. Section 2 gives the details of intrusion detection and its types. Section 3 shows the related work done in this area. Section 4 gives details of proposed framework that uses a meta classifier and dynamic fuzzy clustering algorithm for intrusion detection. Section 5 gives the details of different evaluation criteria through which the performance of an intrusion detection system is measured. Section 6 presents the conclusion and future work.

II. INTRUSION DETECTION

Intrusion is defined as a set of actions that attempt to compromise the integrity, confidentiality or availability of any resource on computing platform [2]. IDS are systems that monitor the network and look for malicious or suspicious behavior in users activity. The main goal of IDS is to alarm the network administrator that any malicious or suspicious activity has happened. According to different detection methods, we can say that there are two types of intrusion detection systems.

A. Misuse Detection

Misuse detection systems are the approach that tries to match user activity to signatures of known attacks that are stored in the database. Such detection systems use a prior defined knowledge to check the new activity happening on the network. It has high speed of detection and low percentage of false alarm. However, it fails in discovering the new attacks that are not defined in the database.

B. Anomaly Detection

Other detection method is the Anomaly Detection approach and it works on principle anomalies are not normal. Such detection approach tries to find whether the change from the normal usage pattern can be called as intrusion or not. Thus, the anomaly detection technique stores the systems normal activity and raises an alarm if any abnormal behavior (i.e. intrusive activity) occurs which differs from normal behavior. Anomaly detection helps in finding new attacks in the network.

This type of IDS can be further divided into two categories: Host-based Intrusion Detection System (HIDS) and Networkbased Intrusion Detection System (NIDS). An HIDS resides on a particular host and looks for indications of attack on that host. An NIDS resides on separate machines that look for indications of attack in the whole network.

III. RELATED WORK

In this section, a study of different fuzzy clustering techniques that is used for developing intrusion detection system is given.

W. Ren [3] has developed a method that applies fuzzy c-means clustering algorithm to detect network intrusion. He carries out fuzzy partition and clustering of data which separates normal data and attack data effectively. His experiment shows the feasibility and validity of fuzzy c-means clustering algorithm.

E. Narayan, et al [4] have proposed algorithms on expectation maximization fuzzy c-means clustering (EMFCM). Proposed algorithms provide better result to fuzzy c-means clustering by avoiding the looping problems and saves time. EMFCM clustering algorithm has fast convergence in a few iterations regardless of the initial number of clusters.

H. Wang [5] has proposed a new hybrid fuzzy clustering algorithm that uses Quantum-behaved Particle Swarm Optimization (QPSO) algorithm and combines with fuzzy c-means (FCM) for abnormally detection. This technique avoids the local minimum problems of FCM by including in the algorithm a strong global searching capacity.

F. Guorui [6] in his paper developed a semi-supervised learning algorithm for intrusion detection which is combined with the fuzzy c-Means algorithm. The KDD CUP99 data set is adopted as the experimental subject.

T. Fries [7] in his paper presents a fuzzy-genetic approach to intrusion detection that is shown to provide performance superior to other GA-based algorithms. In addition, the method demonstrates improved robustness in comparison to other GA-based techniques.

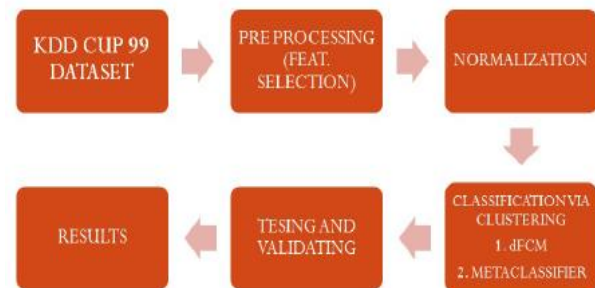
S. Chittineni, et al [8] proposed fuzzy c-means algorithm using neural network algorithm. The proposed work involves two steps. First, an Enhanced K-means Fast Learning Artificial Neural Network

(KFLANN) frame work is used to determine cluster centers. Secondly, fuzzy c-means uses these cluster centers to generate fuzzy membership functions.

Li Jian-guo, et al [9] proposed an improved weighted fuzzy clustering algorithm based on rough set by using the methods of attributes contracted in the rough set theory to improve the FCM algorithm.

IV. FRAMEWORK OF PROPOSED MODEL

The proposed model consists of meta classifier and dynamic fuzzy clustering technique to iteratively do the process of constructing and evaluating detection models. It will make clusters with normal and anomalous traffic in the training dataset. The training dataset and the selected features will serve as the input to the fuzzy clustering algorithm. Finally, the dataset is classified by the metaclassifier.



A. Dataset Description

The data set used was the KDD Cup 1999 Data [13] and it contains a wide variety of intrusions simulated in a military network environment. It consists of approximately 4,900,000 data instances, each of which is a connection record obtained from the raw network data gathered during the simulated intrusions. The attacks come under one of the following four categories: DOS - Denial of Service (e.g. a syn flood), R2L - Remote to Local - Unauthorized access from a remote machine (e.g. password guessing), U2R - User to Root - Unauthorized access to superuser or root functions (e.g. a buffer overflow attack), Probing - surveillance and other probing for vulnerabilities (e.g. port scanning).

B. Weka

WEKA (Waikato Environment for Knowledge Analysis) based on JAVA environment is a free, non-commercial and open-source platform aiming at machine learning and data mining. In WEKA, many popular data mining algorithms are implemented. Users can call the appropriate algorithms according to their needs. However, the FCM and dFCM algorithms are not integrated into WEKA. We intend to implement the two fuzzy algorithms in WEKA.

C. Feature Selection

Feature selection is used commonly in data mining. They are techniques used reducing inputs size for processing and analysis. Feature Selection is important because dataset contain far more information than is

needed to build the model. Feature selection is necessary because the input data can contain redundant or unnecessary information in it. So for selecting the most relevant features from the dataset feature selection is done. Feature selection is always done before the model is trained. Many feature selection techniques are available in WEKA.

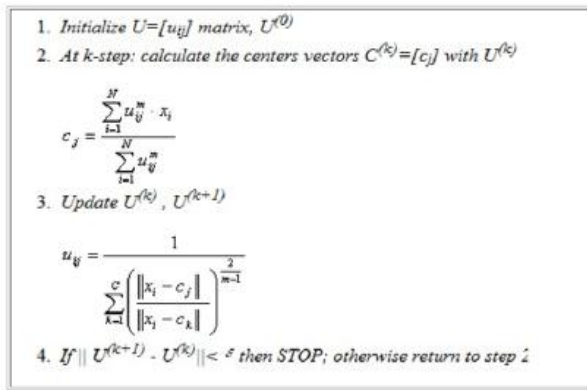
D. Normalization

Normalization technique normalizes all numeric values in the given data. A problem with a typical dataset is that different features are on different scales. So to bring them all in the same range, normalization technique is used. Generally, the resulting values are kept in [0,1] for the data used to compute the normalization intervals.

E. Classification via Clustering

This is a classification technique in WEKA that internally uses a clusterer for classification. Here, firstly the input data is clustered and then the meta-classifier classifies the data into different classes. Internally, any clustering technique can be used to cluster the data. In this paper, dynamic version of fuzzy c means clustering algorithm i.e. dFCM algorithm is proposed. This dFCM algorithm will be implemented and used as the clusterer in WEKA.

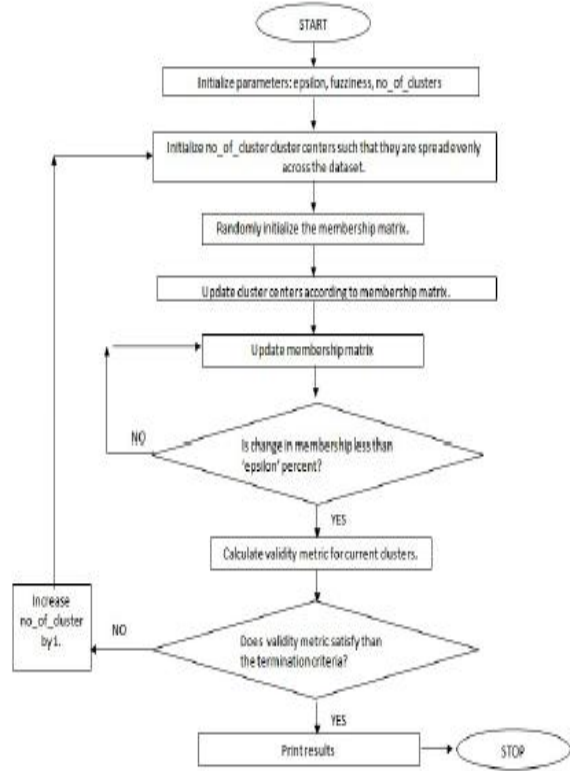
Fuzzy Clustering Algorithm: The main idea in fuzzy clustering is the non-unique partitioning of the data in a form of clusters. The data points are assigned membership values for each of the clusters. The fuzzy clustering algorithms allow the clusters to grow naturally. Fuzzy c means clustering is the most widely used clustering algorithm. The steps of fuzzy c means clustering is given below.



F. Dynamic Fuzzy Clustering Algorithm

The dFCM is a modification of the fuzzy c means algorithm, it allows the cluster centers to be adaptively updated as the data points keep streaming in. If a new cluster is formed, then a new cluster center is automatically generated. The initial steps of dFCM algorithm is similar to the traditional FCM algorithm.

To make it dynamic, cluster validity indexing is done. The working of the dFCM algorithm is shown in the flowchart.



Validity Indices: A validity index is used to determine how well the data is represented by the selected clusters. There are a number of validity indices available. Here, we have used Partition Coefficient index for validity indexing. Partition Coefficient index is based on some of the aspects of the input data. The partition coefficient depends only on the memberships of the data.

V. EVALUATION CRITERIA

Intrusion detection is the process of monitoring and analyzing events that occur in a computer or networked computer system to detect behavior of users that conflict with the intended use of the system. When referring to the performance of IDSs, the following terms are often used when discussing their capabilities:

True positive (TP): classifying an intrusion as an intrusion. The true positive rate is synonymous with detection rate, sensitivity and recall, which are other terms often used in the literature.

False positive (FP): incorrectly classifying normal data as an intrusion. Also known as a false alarm.

True negative (TN): correctly classifying normal data as normal. The true negative rate is also referred to as specificity.

False negative (FN): incorrectly classifying an intrusion as normal.

VI. CONCLUSION AND FUTURE WORK

The dynamic fuzzy c-means clustering technique proposed has a wide range of useful applications. It allows real time cluster analysis by the evaluation of

validity indices when a specified condition is met. This allows a good deal of flexibility within the proposed algorithm, so that it may tackle various types of problems. Other conditions can easily be incorporated into the different phases in the algorithm, depending on the application.

Combination of clustering and classification has been used for model generation. In future for model generation, other clustering and classifiers can be used to improve the detection rate of intrusion detection systems. Experiments can be carried out on different datasets. And also multiclassifier can be used to improve the performance of intrusion detection systems.

REFERENCES

- [1] D.E. Denning, An Intrusion Detection Model, IEEE Transaction on Software Engineering, vol SE-13, no.2, pp. 222-232, 1987.
- [2] K. Labib, Computer Security and Intrusion Detection, from Crosswords, The ACM Students Magazine.
- [3] W. Ren, Application of Network Intrusion Detection Based on Fuzzy C-Means Clustering Algorithm, Intelligent Information Technology Application, 2009. IITA IEEE 2009.
- [4] E. Narayan, P. Singh and G. Tak, Intrusion Detection System Using Fuzzy C Means Clustering with Unsupervised Learning via EM Algorithms, VSRD-IJCSIT, Vol. 2 (6), 2012, 502-510.
- [5] H. Wang, Network intrusion detection based on hybrid Fuzzy C-mean clustering, Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International IEEE Conference.
- [6] F. Guorui, Intrusion detection based on the semi-supervised Fuzzy CMeans clustering algorithm, Consumer Electronics, Communications and Networks (CECNet), 2012 2nd IEEE International Conference.
- [7] T. Fries, A Fuzzy-Genetic Approach to Network Intrusion Detection, Department of Computer Science Coastal Carolina University Conway, South Carolina.
- [8] S. Chittineni and Dr. R.B. Bhogapathi, Neural Network Based Fuzzy C MEANS Clustering Algorithm, International Journal of Electronics Signals and Systems, Vol. 1, Issue 1.
- [9] LI Jian-guo and G. Jing-Wei, Research on Improved Weighted Fuzzy Clustering Algorithm based on Rough Set, Proceedings of International Conference on Computer Engineering and Technology, pp.98- 102, 2009.
- [10] Yu-Ping Zhou, Research on Neuro-fuzzy Inference System in Hierarchical Intrusion Detection, Information Technology and Computer Science, 2009. ITCS 2009 International IEEE Conference.
- [11] M. Panda and M. R. Patra, Network Intrusion Detection Using Naive Bayes. IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.
- [12] B. Thomas and G. Raju, A Novel Fuzzy Clustering Method for Outlier Detection in Data Mining, International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009.

