Volume 5, Issue 5, MAY 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

An Improved Approach for Detecting Unknown Attacks Using Feature Extraction Scheme and Fuzzy-Neural Networks

Vikas Belwal*, Sandip Mandal

Department of Information Technology

DIT University Dehradun, India

Abstract— Internet has been the driving force behind all the technical evolution in the past two decades. It has become the basic requirement of people these days all around the world because of its involvement in banking, e-commerce, e-government etc. Hence, internet security has become a major point of concern for computer scientists. Intrusion Detection System (IDS) is the most efficient tool that is used for internet security. Therefore, Intrusion Detection System is the most discussed subject for computer scientist and researchers these days. Over the period of time various techniques have been proposed and implemented towards this but still lot is needed to be done to make it more efficient. In this paper we are proposing a novel approach for detection of unknown attacks which combines Feature Extraction Scheme and Fuzzy-Neural Networks (FNN) along with K-means Clustering and Support Vector Machines (SVM).

Keywords— Intrusion Detection System; Feature Extraction Scheme; Fuzzy Neural Networks; Support Vector Machine classifier; K-means Clustering.

I. INTRODUCTION

The wide spread use of internet in today's society, especially the exponential rise in importance of e-commerce to the world economy, has made internet security a major international concern. And as we all know it is not technically possible to design a system with no flaws and vulnerabilities at all, intrusion detection has become an important area of research for computer scientist around the world. The concept of the Intrusion Detection System (IDS) was first brought into existence by James Anderson[1] in 1980 and since then IDS has evolved over three decades. Intrusion detection can be defined as an act of identifying or detecting malicious activity which could a threat for the system‟s confidentiality, integrity and availability, and systems that perform such tasks are termed as Intrusion Detection Systems. IDSs can primarily be categorized into two main types: Signature-based i.e. misuse detection and Anomaly-based i.e. anomaly detection IDSs. Signature-based systems are widely used for detecting known attacks but as it is impossible to create signature for attacks before they actually occurred, they are vulnerable to unknown attacks. Whereas Anomaly based IDSs are used for detecting unknown attacks. But they too have drawback of having low detection rate and high false positive rate. To overcome these limitations many researchers have continuously argued for developing an improved IDS which not only be able to detect unknown attacks but also have a least false positive rate.

In order to achieve the set target many impressive techniques have evolved in previous years and it has been widely accepted that data mining concepts can play a crucial role in achieving such targets. In this paper we have proposed a novel approach using the feature extraction scheme, k-means clustering, Fuzzy-Neural Networks (FNN) [2] and Support Vector Machines (SVM) [3] which can be implemented in modern IDSs to reduce the false positive rate to an extent and thus improve the efficiency of the IDSs.

II. RELATED WORK

In last two decades various techniques have been proposed and implemented in IDSs to detect intrusions. At the initial stage rule-based expert systems and statistical approaches were the hot topics for research in the field of intrusion detection. Rule-based expert systems are quite effective in detecting some well-known intrusions with high detection rate but when it comes to novel attacks its vulnerability to such attacks if often exposed, as it is not possible to create signatures for such attacks before they even occurred. On the other hand Statistical-based IDS requires a large amount of data to build mathematical model which is quite complicated and impractical.

To overcome the limitations of above discussed approaches, many researchers proposed various other techniques. Song et al. [4] proposed a different technique, in which the system detects anomalies from the traffic and based on these anomalies statistical features such source address, source port, destination address, destination port, detection time and signature name are extracted. They then apply one-class SVM to these features and detect unknown attacks from signature based IDS alerts. Later Sato et al. [5] modified the previous approach by using more features such as duration, source byte and destination byte.

There were other researchers also who were busy in studying how data mining techniques can be useful, and found that there are various techniques which can be crucial in improving the IDSs. Among these techniques Artificial Neural Networks and Fuzzy logics are considered to be very helpful. Gang Wang [6] in his approach has made use of ANN and

fuzzy clustering. But since both ANN and fuzzy systems have some disadvantages researchers then prepared a model which uses the advantages of these and discards each other's disadvantages to a great extent and they named this model Fuzzy-Neural Network or Neuro-Fuzzy systems. Chandrasekhar [7] in his paper has proposed how fuzzy logics and neural networks can be amalgamated to improve the efficiency of IDSs. Here, in this paper we propose a technique which is a fusion of feature extraction scheme, k-means, FNN and SVM classifier to improve the detection rate and lower the false positive rate.

III. RESEARCH METHODOLOGY

In this section we present the whole framework of the proposed methodology. Then we have elaborated the four modules i.e., Feature extraction scheme, K-means clustering, Fuzzy-Neural networks and SVM Classifier in brief.

The proposed method consists of the following 4 steps:

- Step 1: Feature extractions.
- Step 2: K-means clustering.
- Step 3: Fuzzy-neural networks.
- Step 4: SVM classifier.

The block diagram our approach is given in Fig.1.

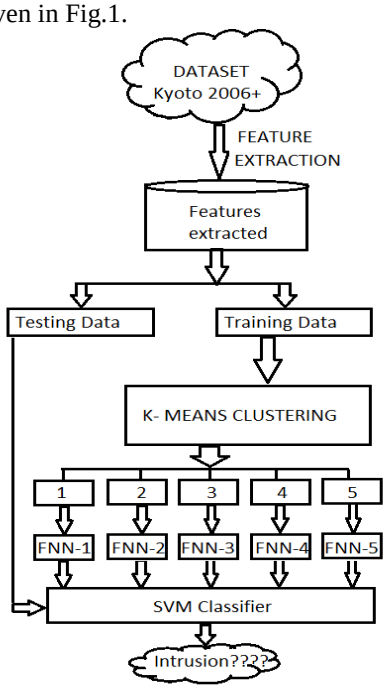

Fig1. Block diagram of the proposed method

A. Feature Extraction

This module aims at analysing the network traffic and then the extraction of various features of packets or data travelling in the network based on the IDS alerts. We will be using Kyoto2006+ [8] dataset for this.

The sample of features extracted is shown in Table 1.

TABLE I. SAMPLE OF EXTRACTED FEATURES

Duration	Src_byte	Dst_byte	Src_IP	Dst_IP
3.03	2137	197	2d.x.x.0	8d.x.x.12
66.01	0	0	2d.x.x.0	8d.x.x.10
2.42	520	1879	2d.x.x.0	8d.x.x.22
0.12	192	192	2d.x.x.0	8d.x.x.31

B. K-Means Clustering

K-means clustering is a prototype based clustering technique that attempts user defined number (k) of clusters, which are represented by their centroids. K-means is one of the easiest unsupervised clustering algorithm that solve the well known problems in different fields. The k-means algorithm takes the input character 'k' and partitions a set of 'n' data

points into k clusters so that the resulting intra-cluster similarity is high but the inter cluster similarity is low. The aim of using k-means cluster is to partition a given set of data into clusters, where data belonging to different cluster should be as different as possible.

C. Fuzzy Neural Networks

Neural Network is an efficient tool for classification. The high level of tolerance makes neural networks flexible and efficient in IDS. Neural networks are very good at recognizing patterns. But it has many disadvantages too, such as, of having impossible interpretation of the functionality and also faces difficulty in estimating the number of layers and number of neurons. These disadvantages can be overcome by combining fuzzy into neural networks and consequences in better results and outcomes. A neuro-fuzzy system is a fuzzy system that uses a learning algorithm derived from or inspired by neural network theory to determine its parameters by processing data samples. The major advantages of using neuro-fuzzy are that it can handle any kind of information (numeric, linguistic, logical etc). It can manage imprecise, partial value or half-perfect information. It can resolve conflicts by collaboration and aggregation. It has self-learning, self-organizing and self-tuning capabilities. There is no need of prior knowledge of relationships of data as in case of human decision making process. It can perform fast computation using fuzzy number operations.

D. Support Vector Machine(SVM) Classifier

The support vector machine (SVM) is a supervised classification system that minimizes an upper bound on its expected error. It attempts to find the hyperplane separating two classes of data that will generalize best to future data. Such a hyperplane is the so called maximum margin hyperplane, which maximizes the distance to the closest points from each class. Generally, they work well when the number of features is magnitudes higher than the available training data. They also avoid the two problems of dimensionality; they generalize well to unseen data and they are efficient as they avoid explicit use of higher order dimensional spaces. SVM classifier is used as it produces better results for binary classification when compared with other classifiers. The data with constrained number of attributes is directed to the SVM which is binary, classified to detect if there is any intrusion or not.

IV. CONCLUSIONS

Here, in this paper we have proposed a novel approach for detecting unknown attacks using various data mining concepts and also discussed the different modules of our approach in brief. The approach we are proposing in this paper is expected to be very efficient as we have combined the various data mining techniques and this not only multiply their strength but also delimits each other limitation. It can also be concluded that though a lot has been improved in IDSs since it was first developed two decades back, still a lot has to be done and data mining concepts could play a crucial role in improving its efficiency.

REFERENCES

[1] J.P.Anderson, "Computer security threat monitoring and surveillance," James P. Anderson Co., Washington, 1980.

[2] R.Jang, "Neuro-Fuzzy modelling: Architecture, analysis and Application," Ph.D Thesis, University of California, Berkley,1992.

[3] S.Mukkamala, G.I.Janoski, A.H.Sung, "Intrusion Detection Using Support Vector Machines," in proceedings of the High Performance Computing Symposium- HPC, San Diego, CA, USA,2002, pp.178-183.

[4] J. Song, H. Takakura, and Y. Kwon, "A generalized feature extraction scheme to detect 0-day attacks via ids alerts," in Applications and the Internet, 2008. SAINT 2008. International Symposium on. IEEE, 2008, pp. 55–61.

[5] M.Sato, H.Yamaki, H.Takakura, "Unknown Attacks Detection Using Feature Extraction from Anomaly-based IDS Alerts," 12th International Symposium on Applications and the Internet, IEEE/IPSJ, 2012.

[6] Gang Wang, Jinxing Hao, Jian Ma, Lihua Huang, "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering," Expert Systems with Applications, http://www.elsevier.com/locate/eswa/.

[7] A.M. Chandrasekhar, K.Raghuveer, "Intrusion Detection Technique using k-means, Fuzzy Neural Networks and SVM Classifiers," in International Conference on Computer Communication and Informatics(ICCCI), IEEE, 2013.

[8] "Kyoto2006+ dataset," http://www.takakura.com/Kyoto data/.

