SOUTHEAST EUROPE JOURNAL OF SOFT COMPUTING
Available online at www.scjournal.com.ba

Intrusion Detection System using Fuzzy Logic

Alma Husagic-Selman

International University of Sarajevo, Faculty of Engineering and Natural Sciences
Hrasnicka Cesta 15, 71000 Sarajevo, Bosnia and Herzegovina
Ahusagic.selman@gmail.com

Abstract

Intrusion detection plays an important role in today's computer and communication technology. As such it is very important to design time efficient Intrusion Detection System (IDS) low in both, False Positive Rate (FPR) and False Negative Rate (FNR), but high in attack detection precision. To achieve that, this paper proposes IDS model based on Fuzzy Logic. Proposed model consists of three parts, Input Reduction System (IRS), which uses Principal Component Analysis to reduce the dimensions of the system from 41 to 10, Classification System, which uses Fuzzy C Means to create data clusters based on training data and Pattern Recognition System based on Nearest Neighborhood method, which classifies new-coming data records to their respective clusters. Based on different attack types, the system performance in classification process is different and the best performance is achieved for PROBE attack, with 99.3% success rate, and the best performance in pattern recognition is achieved for U2R with 58.8% of success rate.

Keywords: Intrusion detection, Intelligent Intrusion Detection System, Fuzzy C Means, Nearest Neighborhood.

1. Introduction

Since second half of last century, computer networks started to grow with tremendous speed and with them the need for security mechanisms which would ensure data, privacy and computer security grew as well. Many different security mechanisms were designed, yet none was reliable enough to protect the computer-network system from ever evolving threats and attacks. Firewalls were made in order to protect the networks from attacks that come from outside world, but they do not obliterate any intrusion coming from inside the network. Intrusion Detection Systems (IDS), on the other hand, monitor networking packets in order to prevent any form of computer attacks from within the network [1-4]. This work focuses on IDS, since existing commercial IDS's offer wide window for improvements.

In general, IDSs may be designed to perform misuse detection or anomaly detection [1, 5]. In misuse detection, all known abnormal behavior is defined and the system is trained to recognize it. It works by comparing arriving packet with features of known attack behavior. If any new, not predefined attack arrives, the system would recognize it as normal packet, causing high FNR [2].To avoid very high FNR, misuse based IDS must be retrained very often, sometimes causing delays in the network [6].

Anomaly detection is modeled based on normal behavior [7], so any pattern violating that behavior would be defined as system attack [1, 5]. Anomaly detection causes high FPR, because even new normal packet, unknown to the system, would be identified as an attack. This deteriorates overall network performance, since some normal packets would never reach destination. For these reasons, most commercial IDS are designed to perform misuse detection alone.

False alarms, be it false positive or false negative, are limiting the performance of IDS. It is therefore very important to reduce both types of these alarms, and the best way to do it is by combining anomaly and misuse detection [5, 8].

This paper proposes fast and efficient IDS based on fuzzy logic, that will first train the system to cluster the data into different clusters and then each new-coming packet would be classified using pattern recognition into an appropriate cluster. The proposed model produces five outputs, normal packet and four attack types (DoS, U2R, R2L and Probe) [9]. The model proposed here was tested in different ways, and that will be shown in the last two sections of this paper.

The paper is organized as follows: Related works are discussed in Section 2, Section 3 gives an overview of methods and algorithms used in this work, Section 4 presents data used for experimentation, Section 5 describes the system model, Section 6 presents and discusses results, and Section 7 concludes the paper.

2. Related Works

Different machine learning mechanisms, including Artificial Neural Networks, Fuzzy Logic, Genetic Algorithms, etc. have been used on KDD CUP 1999 data for Intrusion Detection [1-10], with neural networks as main tool in this type of problem. Different neural network algorithms have been used, including Grey Neural Networks [4], RBF [10,11] Recirculation Neural Networks [2], PCA [6, 12] and MLP[5], with MLP generally showing better results than others [2].These works are mainly focusing on misuse detection. In order to combine misuse and anomaly detection, many researchers have recently attempted hybrid methods, by combining neural networks with other machine learning mechanisms, such as fuzzy logic or genetic algorithms [5, 13-16]. Fuzzy logic tends to be better tool of clustering, as it is faster and more suitable for real-time systems. Summary of results based on intelligent methods is presented in Table 1.

Table 1: Summary of results for different neural networks and hybrid systems

Approach	DOS, %	Probe, %	R2L, %	U2R, %
Decision Trees	99.80	50.00	33.30	50.00
Bayesian Networks	99.70	52.60	46.20	25.00
Flexible Neural Tree	98.80	99.30	98.80	99.90
Fuzzy NN	100.00	100.00	99.80	40.00
MLP	99.90	48.10	93.20	83.30
Advanced NN	98.97	94.62	97.02	59.00
Evolving Fuzzy NN	98.99	99.88	97.26	65.00
Recirculation NN	97.89	98.15	98.22	100.00
PCA & Gray NN	68.00	88.00	58.00	26.00
Fuzzy C-Mean and MLP	100.00	99.80	40.00	100.00
ANFIS, FIS & GA	99.70	84.97	31.68	16.67

Table 2 shows FPR and FNR for different artificial intelligence classification algorithms.

Table 2: Summary of FPR and FNR for different classification algorithms

Approach	FNR, %	FPR, %
Flexible Neural Tree	1.2	0.3
MLP	5.8	0.8
Clasterization	7	10
K-NN	9	8
SVM	2	10
Recirculation Neural Networks	1.83	0.03
Fuzzy C-Mean and MLP	0.01	0.01

Table 2 shows that Fuzzy C-Means combined with MLP ANN has the lowest FPN and FNR, and as such this paper tries to examine the performance of pure fuzzy logic-based IDS.

3. Algorithms and Methods

Multiple methods are used in this work: PCA for feature reduction, Fuzzy C Means for data classification, and Nearest Neighborhood Method for pattern recognition.

3.1 Principle Component Analysis

PCA is very useful mathematical algorithm, based on orthogonal linear transformation, which is widely used for data compression, image processing and feature extraction [6, 12]. The goal of PCA is to find a set of orthogonal components that minimize the error in reconstructed data. An equivalent formulation of PCA is to find an orthogonal set of vectors that maximize the variance of the projected data [17]. In other words, PCA transforms the data into different frame of reference with minimal error and using fewer features than the original data, while preserving data randomness. [18]. For more detailed description of PCA algorithm refer to [17, 18].

3.2 Data Clustering and Classification

Data clustering and classification is the process of creation of clusters given the initial training data. These clusters can then be used in combination with other methods, such as neural networks, or fuzzy pattern recognition. The classification method used in this paper is called Fuzzy C-Means, and it is a method of clustering based on minimization of objective function J_m.

$$J_m(U, v) = \sum_{i=1}^{N} \sum_{j=1}^{C} (\mu_{ij})^{m'} (d_{ij})^2 \text{ , } 1 \le m \le \infty \qquad (1)$$

where U is partition matrix, v_i is cluster center, d_{ij} is Eucledian distance measure in m-dimensional feature space, between the j^{th} data sample x_j and the i^{th} cluster center v_i, and μ_{ij} is the membership of j^{th} data point to the i^{th} class.

Partition matrix U is used for grouping a collection of n data sets into c classes, and as such each entry in the partition matrix is represented by the membership function μ_{ij}. The Eucledian distance and cluster centers are given in equations (2) and (3).

$$d_{ij} = \left[\sum_{k=1}^{m} \left(x_{jk} - v_{ik} \right)^2 \right]^{1/2} \qquad (2)$$

$$v_{ik} = \frac{\sum_{j=1}^{n} \mu_{ij}^{m'} * x_{ji}}{\sum_{j=1}^{n} \mu_{ij}^{m'}} \qquad (3)$$

k is a variable on the feature space and m' is the membership exponent which controls the level of fuzziness.

The fuzzy C means is trying to tune the partition matrix, centers and distances, so that the objective function J_m is minimized [26].

3.3 Pattern Recognition

Pattern recognition is defined as a process of identifying structure in data by comparing is to some known structure, generally developed through methods of classification, such as Fuzzy C means. Multiple methods for pattern recognition exist, and in this research work we focus on Nearest Neighborhood Method, which is suitable for multi-feature pattern recognition process.

In the nearest neighbor classifier, m features for each data sample is considered as a vector,

$$x_i = \{ x_{i1}, x_{i2}, x_{i3}, \ldots, x_{i\,m} \} \qquad (4)$$

Assuming that the clusters already exist, then the incoming data samples can be classified to their respective clusters by calculating the distance d between the data sample and the center of each cluster. The data sample x will then be classified to belong to the cluster to which center it has the shortest distance, as shown in the equation 5 [26].

$$d\,(x, x_i) = min\{d(x, x_k)\} \quad 1 \le k \le n \qquad (5)$$

4. Data Description

The data used in this work is widely used KDD CUP 1999 data, which was created based on DARPA Intrusion Detection data set, collected by MIT Lincoln Laboratory. [9].

The data contains 41 features, specifying packet type, protocol and so on, and class label, specifying if the packet is normal or attack. Data set contains 22 attack types, which can be divided into four main categories [9], as follows:

-Denial of Service (DoS) denies service to legitimate users, most commonly through overloading of existing resources. Six out of total 22 attack fall into this group.

-User-to-Root (U2R), user with normal user privileges tries to exploit vulnerabilities of the system in order to gain the access to the root of the system. Four out of total 22 attack fall into this group.

-Remote-to-Local (R2L), unauthorized user from a remote machine tries to access local machine by exploiting holes in local machine. Eight out of 22 attacks fall into this group.

-Probing (Probe), unauthorized user monitors the networks in order to obtain information and discover system's vulnerabilities. Four out of total 22 attack fall into this group.

Original KDD CUP 1999 training data, consisting of about 5 million records, was too large to analyze, and for that reason, concise set known as '10% training set' was used. Out of this concise set of 500 000, 4911 data records were selectively chosen to represent the all possible types of packets and were used in training and testing of Fuzzy IDS system presented here.

5. Intrusion Detection Model

Intrusion Detection Model was designed based on anomaly detection and misuse detection, with Fuzzy C Means recognizing if the attack exists or not, and if it exists which attack it is. To reduce the FPR and FNR, there is an update system, which helps update the classification clusters. Thus, proposed Intrusion Detection Model consists of three main parts: Input Reduction System, Classification system and Pattern Recognition System (Figure 1). Pattern Recognition with new data and Classification together represent the update system, or the system responsible for reducing FPR and FNR.

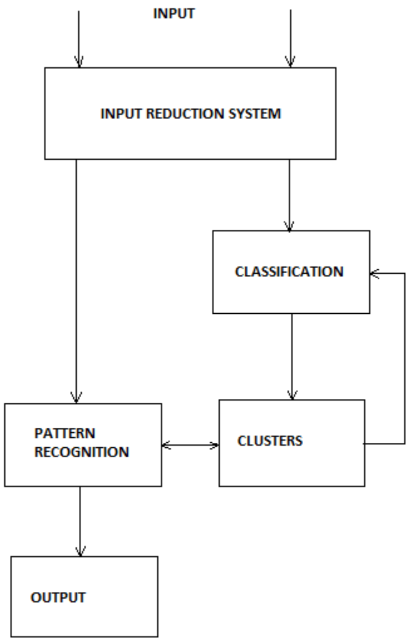

Figure 1: Proposed Intrusion Detection Model

5.1 Input Reduction System

In systems with large dataset characterized with numerous features, input or feature reduction process should be done whenever possible. This step helps remove distracting variance from a dataset and as such improves the performance of the classifier and speeds up the classification process. In this work, single PCA neural network is was chosen as a tool for feature reduction.

PCA Neural Network takes original 41 inputs and reduces the input size to 10. The initial number of PCA was chosen 10, and the system performance was checked accordingly. Then the number of PCA was increased to 15 and then to 41, and the overall performance of the algorithm did not improve. Rather it deteriorated. As scuh it was decided to keep the initial number of PCA components, i.e. altogether 10. General overview of input reduction system is shown in Figure 4.

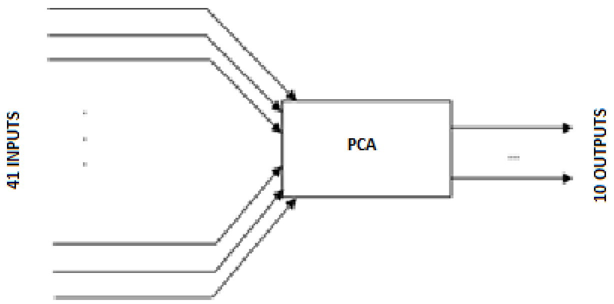

Figure 2: Input/Feature Reduction System

5.2 Clustering based on Fuzzy C Means

Fuzzy Classification system is based on Fuzzy C Means, which receives inputs from IRS, with each input having 10 features. The Fuzzy C Means was done to create 2 clusters, and 5 clusters. Five-cluster system is shown in Figure 5, and it represents the fuzzy system able to recognize all 5 types of packets, namely normal, dos probe, u2r and r2l.

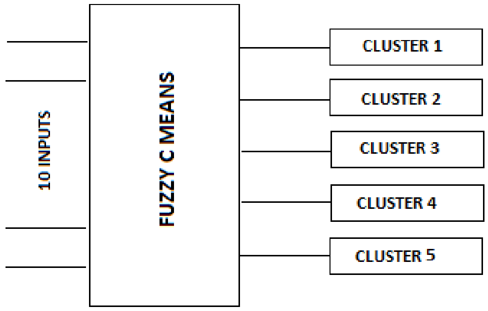

Figure 3: Fuzzy C Means clustering scheme

To test the success of the performance when the number of cluster is reduced, four two-cluster systems were also simulated. The four two cluster systems were separately trained to recognize the following four sets: normal packet vs. DOS attack, normal packet vs. PROBE attack, normal packet vs. U2R attack and normal packet vs. R2L attack. The further discussion on the testing criteria, training and test data as well as the performance of each of these systems is discussed in the following section.

The clusters and their centers were then used as base for the process of pattern recognition.

5.3 Pattern Recognition

Pattern recognition system was based on Nearest Neighborhood method, which relies on previously generated clusters. Each input data is evaluated based on each feature and the distance of each feature and corresponding data center is calculated. The system then classifies the input to belong to that cluster to which it has the shortest distance.

6. Results and Discussion

The Fuzzy IDS was simulated and tested via two different approaches, one was to cluster and recognize all types of packets, and the other was to cluster and recognize four pairs having pattern “normal packet vs. some attack type”. This was done to determine the success in classification of each type of attack using proposed model.

The simulation of each approach was done as follows: first the training and test input data underwent the reduction process via PCA, so that the features of the input data were reduced from 41 to 10. The training data was then partitioned into initial clusters, and mean of each cluster, together with initial partition matrix U was calculated. The process then underwent the modification of centers, until the difference between the new center and the old one was 10^{-8}. The values of partition matrix and new clusters were then recorded and these values were used as base for pattern recognition process.

Pattern recognition used PCA reduced test data. During that process the difference between the input data record and all cluster centers was determined, and smallest difference determined to which cluster that data record belongs to. The error of this classification is then calculated and results are shown in Table 3.

Table 3: Successful Classification using Fuzzy C means and Nearest Neighborhood method.

	All Attacks	Normal vs. DOS	Normal vs. PROBE	Normal vs. U2R	Normal vs. R2L
Fuzzy C-Means	20%	51.8%	99.33%	69.1%	58.7%
Nearest Neighborhood	35.7%	53%	51.7%	58.8%	45%

When all types of packets were used, i.e. normal packets and all five attacks, the size of training data was 500, and the size of test data was 280. The Fuzzy C means created five clusters, and the success rate was 20%. This low success rate can be attributed to the nature of data records, which has high overlapping.

The best performance in Fuzzy C means was detected when the system was trained to generate two clusters, one for normal packets and one for PROBE attack, the performance success was 99.3%.

The best performance in Pattern Recognition process, amounting to 58.8 %, was achieved with two clusters and U2R attack type. For this type of attack, the classification success was 69.1%.

These results will be improved for larger input training data sets, but due to computational limit of the devices used in this simulation, this was not done.

7. Conclusion and Recommendation

In this paper we have proposed new model for Intrusion Detection System. The model consists of three parts: Input reduction system, classification system and pattern recognition system.

Input reduction system is based on PCA, and it reduces the number of inputs from 41 to 13. Classification system is based on Fuzzy C Means and it produces the clusters that will be used in pattern recognition process. The pattern recognition system is used to classify new, unknown data to its corresponding cluster. The pattern recognition process proposed in the model should send the feedback to the classification system in order to update the centers if new, untrained data comes.

The overall system showed the classification for five clusters was 35.7 %, and the best overall performance was achieved for U2R attack with 58.8 % correctness. The classification system's best performance was 99.3% for normal packet and DOS attack.

The future improvements on this research could be done in the respect that the training data size increases, so that the cluster centers get more tuning. And another way to improve the system was to include neural network committee machine after the clustering process.

References

[1] Chavan, S.; Shah, K.; Dave, N.; Mukherjee, S.; Abraham, A.; Sanyal, S.; , "Adaptive neuro-fuzzy intrusion detection systems," *Proceedings. ITCC 2004. International Conference on Information Technology: Coding and Computing, 2004.*, vol.1, no., pp. 70- 74 Vol.1, 5-7 April 2004.

[2] P. Kachurka, V.Golovko, "Neural Network Approach to Real-Time Network Intrusion Detection and Recognition". *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, Prague,15-17 September 2011.

[3] Jinguo Zhao; Min Chen; QinyunLuo; , "Research of intrusion detection system based on neural networks," *IEEE 3rd International Conference on Communication Software and Networks (ICCSN), 2011*, vol., no., pp.174-178, 27-29 May 2011

[4] Dong-Xue Xia; Shu-Hong Yang; Chun-Gui Li; , "Intrusion Detection System Based on Principal Component Analysis and Grey Neural Networks," *2010 Second International Conference on Networks Security Wireless Communications and Trusted Computing (NSWCTC)*, , vol.2, no., pp.142-145, 24-25 April 2010

[5] M.J. Muna, M. Mehrotra, "Design network intrusion detection system using hybrid fuzzy-neural network," *International Journal of Computer Science and Security*. 2010. vol. 4 (3). pp. 258–294.

[6] Fengxi Song; ZhongweiGuo; Dayong Mei; , "Feature Selection Using Principal Component Analysis,", *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM)*, vol.1, no., pp.27-30, 12-14 Nov. 2010.

[7] Dae-Ki Kang; Fuller, D.; Honavar, V.; , "Learning classifiers for misuse and anomaly detection using a bag of system calls representation," *Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop, 2005. IAW '05.*, vol., no., pp. 118-125, 15-17 June 2005.

[8] G. Wang, J. Hao, J. Ma, L. Huang, "A new approach to intrusion detection using artificial neural networks and fuzzy clastering, "*Expert Systems with Applications*. Issue 2010.No 2.

[9] KDD Cup’99 Competition, 1999, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[10] C. Zhang, J. Jiang, M. Kamel, “Intrusion detection using hierarchical neural networks*”. Elsevier - Pattern Recognition Letters*, Volume 26, Issue 6, 1 May 2005, Pages 779–791

[11] Ahmed, U.; Masood, A.; , "Host based intrusion detection using RBF neural networks," *Emerging Technologies, 2009. ICET 2009. International Conference on* , vol., no., pp.48-51, 19-20 Oct. 2009.

[12] ShilpaLakhinaet. al. “Feature Reduction using Principal Component Analysis for Effective Anomaly–Based Intrusion Detection on NSL-KDD”, *International Journal of Engineering Science and Technology*, Vol. 2(6), 2010, 1790-1799

[13] Fan Li; , "Hybrid Neural Network Intrusion Detection System Using Genetic Algorithm," *International Conference on Multimedia Technology (ICMT), 2010*, vol., no., pp.1-4, 29-31 Oct. 2010.

[14] A. N. Toosi, M. Kahani, “A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers”. *Elsevier-Computer Communications* 30 (2007) 2201–2212.

[15] S. Mukkamala, G.Janoski, A. Sung, “Intrusion Detection Using Neural Networks and Support Vector Machines”. *Proceedings of 2002 IEEE International Joint Conference on Neural Networks (IJCNN '02)*, Honolulu, May 12-17, 2002.

[16] S. Chavan, K. Shah, N. Dave and S. Mukherjee, “Adaptive Neuro-Fuzzy Intrusion Detection Systems". *Proceedings of IEEE International Conference on Information Technology: Coding and Computing ITCC*, 2004.

[17] Diamantaras, K. I. and Kung, S. Y. (1996).”Principal Component Neural Networks". *John Wiley & Sons*, New York.

[18] I.T. Jolliffe, “Principal Component Analysis” Springer-Verlag, 1986.

[19] S. Haykin, “Neural Networks A Comprehensive Foundation”, Second Edition, *Prentice-Hall*, *Inc.* Simon &Schuster, A Viacom Company Upper Saddle River, New Jersey 07458, 1999.

[20] J. Goldsmith, “Unsupervised learning of the morphology of a natural language”, *Computational Linguistics*, Vol. 27(2), pages 153198, 2001.

[21] Suvad.S., Alma H. S., (2011). “Multilayered Feed forward Neural Networks as a Tool for Distinction of the Authors of Texts". *Proceedings of the IEEE XIII International Symposium on Information, Communication and Automation Technologies*, Sarajevo, October 27-29, 2011.

[22] R. E. Schapire, “The strength of weak learnability”, *Machine Learning*, vol.5, pp.197-227, 1990.

[23] Moller, M.F. “A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning”, *Elsevier- Neural Networks*, Volume 6, Issue 4, 1993, Pages 525–533.

[24] N. J. Nilsson, “Learning Machines: Foundations of Trainable Pattern-Classifying Systems”, New York: *McGraw-Hill*, 1965.

[25] M. H. Beale, M. T. Hagan and H. B. Demuth. "Neural Networks Toolbox. User's Guide". Matlab R2012a, MathWorks Inc.

[26] 10. T. Ross, “Fuzzy Logic with Engineering Applications”. John Wiley & Sons, Aug 16, 2004.