Performance Enhancement of Intrusion Detection using Neuro - Fuzzy Intelligent System

Dr. K. S. Anil Kumar
Associate Professor
Department of Computer Science
Sree Ayyappa College Eramallikkara,
Chengannur, Kerala, drksanil@gmail.com

Dr. V. Nanda Mohan
Professor Emeritus in Knowledge Management
Department of Futures Studies
University of Kerala, Thirvananthapuram-34(India)
nandamohanv@gmail.com

Abstract

This research work aims at developing hybrid algorithms using data mining techniques for the effective enhancement of anomaly intrusion detection performance. Many proposed algorithms have not addressed their reliability with varying amount of malicious activity or their adaptability for real time use. The study incorporates a theoretical basis for improvement in performance of IDS using K-medoids Algorithm, Fuzzy Set Algorithm, Fuzzy Rule System and Neural Network techniques. Also statistical significance of estimates has been looked into for finalizing the best one using DARPA network traffic datasets.

Keywords: Intrusion *Detection System (IDS), K-medoids algorithm, Neuro - Fuzzy, DARPA dataset*

1. Introduction

The importance of protecting computer networks and vital information from attacks has become inevitable significance for deploying Intrusion Detection Systems (IDS). Intrusion detection types can be broadly classified into Anomaly detection and Misuse detection. An Anomaly based IDS is a system for detecting computer intrusions and misuse by monitoring system activity and classifying it as either normal or anomalous. The classification is based on heuristics or rules, rather than patterns or signatures, and attempts to detect any type of misuse that falls out of normal system operation [1]. Misuse detection is an approach in detecting attacks. In misuse detection approach, we define abnormal system behavior at first, and then define any other behavior, as normal behavior [2].

The drive for using the hybrid intelligent systems is to improve the accuracy of the intrusion detection inference system when compared to individual system. With the use of single intelligent IDS, it is not possible to achieve an acceptable rate of detection. In view of the enormous computing power available with the present day processors, combining multiple IDS to obtain best-of-breed solutions has been attempted earlier. The hybrid intrusion detection model combines the individual base classifiers and other hybrid machine learning paradigms to maximize detection accuracy and minimize computational complexity [3].

Anomaly based algorithms may miss a large volume of abnormal activity and also trigger false positive alarms from genuine yet rare events. Our approach is to develop a novel anomaly based intrusion detection inference system, which is fast, accurate, efficient, having high detection rate and yielding very low false positives. We have developed a hybrid model – Fuzzy Neural Network Model (FNNM) – Comprising of K-Medoids Algorithm, Fuzzy Sets Algorithm, Fuzzy Rule System and Neural Network techniques.

2. Rational of the Study

This section gives a brief introduction of the techniques used in the proposed work.

2.1 Architecture of the FNNM

The Fuzzy Neural Network Model used various algorithms apt for fulfilling the requirements at every stage. We have used K-Medoids is to cluster normal and intrusion datasets and fuzzy sets from both the clusters using Fuzzy-Set algorithm. Fuzzy Rule Algorithm is used for rule generation from the perceived behavior of normal and abnormal clusters, and Neural Networks using Mat Lab for detecting abnormal packets similar to the ones given during learning sessions and for an artificial intelligence that detects anomalies not presented during learning.

2.2 K-Medoid Clustering Algorithm

To overcome the disadvantages of K-means and to improve detection rate, a new clustering approach, K-medoid algorithm is used. The K-medoid algorithm is a partitioning technique of clusters that clusters the data sets of n objects into K clusters. It could be more robust to noise and outliers as compared to K-means since it minimize a sum of pair wise dissimilarities using a squared Euclidean distance [4]. DARPA dataset is categorized into just two clusters, normal and intrusion and hence the value of K can be simply defined as ‘2’[5,6].

2.3 Fuzzy Set Algorithm

The process of a fuzzy system has three steps. These steps are Fuzzification, Rule Evaluation, and Defuzzification. In the fuzzification step, the input crisp values are transformed into degrees of membership in the fuzzy sets. In the rule evaluation step, each fuzzy rule is assigned with a strength value. The strength is determined by the degrees of memberships of the crisp input values in the fuzzy sets of antecedent part of the fuzzy rule. The defuzzification stage transposes the fuzzy outputs into crisp values [7].

Fuzzy set theory provides a mathematical framework for representing and treating uncertainty, imprecision, and approximate reasoning [8,9]. Fuzzy set operations such as union, intersection, and compliment are similar to those of ordinary set operations. The union of two fuzzy sets A and B is a fuzzy set C, where $C = A \cup B$ and whose membership functions are related by the following equation:

$$\mu C(x) = \max(\mu A(x), \mu B(x)) = \mu A(x) \vee \mu B(x)$$

The intersection of two fuzzy sets A and B is a fuzzy set C, where $C = A \cap B$ and are related by the following equation:

$$\mu X(\xi) = \min(\mu A(x), \mu B(x)) = \mu A(x) \wedge \mu B(x)$$

The complement of two fuzzy sets A and B is a fuzzy set C, whose membership functions are related by the following equation:

$$\mu(\text{not}A(x)) = 1 - (\mu A(x))$$

Gunes Kayacik et al. [10], Herve Debar et al. [11] and McHugh [12] have used Fuzzy set theory which provides a mathematical framework for representing and treating uncertainty, imprecision and approximate reasoning. In this work, we have explored the count and the uniqueness features of the fuzzy system. The pseudo-code used for the Fuzzy intrusion set and non intrusion set are given in Figure 2.3.1

```
 Pseudo code: Fuzzy set algorithm
A=selected attribute         Intrusion/non intrusion cluster
S=subset of operation      Used to store unique values
K=next element from the available data
S=item[i]
For i=1 to n-1
            K=DataFIeld [i+1]
            S=S union K       Select unique item of the field
End for
Store S
```

Figure2.3.1: The pseudo code for the Fuzzy set algorithm

2.4 Fuzzy Rule Algorithm

The fuzzy rule algorithm is developed for creating the training dataset for the FNNM. The SQL queries are developed using the “RandAndOr” function for short-listing the distinct values contained in each field. These values symbolize the characteristics of abnormality in the intrusion data and normality from the normal data. The rule creation generated a logical sequence which contains the “and” and “or” logical operators and impact the decision of abnormality or normality are represented in terms of weights assigned [13].

```
Psuedocode: Fuzzy Rule Algorithm
Initialize Increment to 1
Initialize Weight of Find Record to 0
Initialize Qcnt to 1
WHILE Increment < NI
        FOR each value FL
        Index [FL] = rand() mod Nfl
        ENDFOR
        FOR each value IL
        QStr = sql select statement where each
        Field[IL] = Index[IL] + ‘ ‘ + RandAndOr();
        ENDFOR
        TotFR = ExecuteQuery(Qstr)
        IF TotFR is non zero THEN
        Wht[Qcnt ] = TotR / TotFR
        Add 1 to Qcnt
        ENDIF
        Add 1 to Increment
ENDWHILE
Save Wht
Save Qstr
```

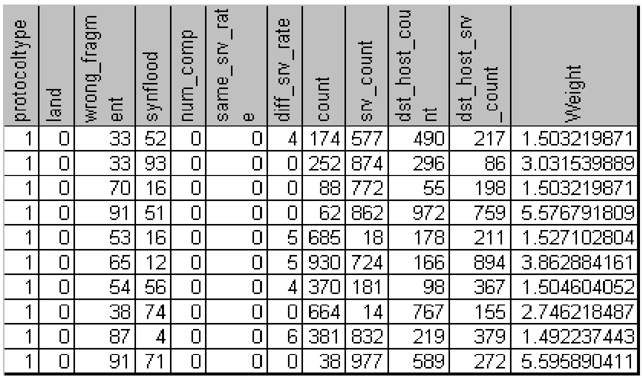
Figure 2.4.1: The pseudo code for the Fuzzy Rule algorithm

3 Methodology

The fields in the DARPA datasets are scrutinized and categorized as intrusion and normal data by applying K-medoid clustering technique. The clustered dataset is given to the fuzzy system for creating Intrusion fuzzy set and Non Intrusion fuzzy set. The detection performance of the Intrusion detection system is very difficult in large volume of network data traffic in real time system. So a lightweight and robust IDS is a must for detecting novel attacks in a real time system. The experimental results proved that the fuzzy set algorithm reduced the size of training data. The time factor for developing fuzzy set is very less and it is in terms of fraction of seconds.

The intrusion fuzzy set and non-intrusion fuzzy set is used to generate fuzzy rules using the Fuzzy Rule algorithm. This algorithm formulates fuzzy rule using the knowledge gained through analysis of instances which helps it to discriminate the Normal and Intrusion fuzzy sets. The fuzzy rule places the entire normal and abnormal data in a separate set or a vector. And hence when a data is received the fuzzy logic itself can classify the regular data from the deviant one.

The values of the data fields present in each packet are temporarily stored in a vector. Using a random function all the possible combinations of these values scattered across the fields is analyzed. The algorithm appraises the values of these parameters along with the measured count of its presence. Then it calibrates the weights in accordance by dividing the counts thus derived by the total number of records present in each cluster. These weights denote the degree to which the presence of a particular attribute has influenced the presence of abnormality of the intrusion data. The resulting weights are stored as the 12th attribute of the record and these weights are subsequently forwarded to the Mat Lab backpropagation algorithm for learning the neural network as shown in Table 3.1.



protocoltype	land	wrong_fragment	synflood	num_comp	same_srv_rate	diff_srv_rate	count	srv_count	dst_host_count	dst_host_srv_count	Weight
1	0	33	52	0	0	4	174	577	490	217	1.503219871
1	0	33	93	0	0	0	252	874	296	86	3.031539889
1	0	70	16	0	0	0	88	772	55	198	1.503219871
1	0	91	51	0	0	0	62	862	972	759	5.576791809
1	0	53	16	0	0	5	685	18	178	211	1.527102804
1	0	65	12	0	0	5	930	724	166	894	3.862884161
1	0	54	56	0	0	4	370	181	98	367	1.504604052
1	0	38	74	0	0	0	664	14	767	155	2.746218487
1	0	87	4	0	0	6	381	832	219	379	1.492237443
1	0	91	71	0	0	0	38	977	589	272	5.595890411

Table 3.1: The data features with computed weights.

4. Experimental Results

The training data set were collected using random sampling method from DARPA dataset of 5000 records. The FNNM offers a very high detection rate of 99.86% and significant reduction in false alarm rate of 0.72%. We have achieved an average value of very high accuracy of 99.17 percent when tested with the different volumes of datasets as shown in table 4.1.

Table 4.1: Analysis of FNNM test results

Analysis	Percent of Result
False Positive Rate	0.72
Probability of Detection	99.86
Overall accuracy	99.17

5. Conclusions

Benchmarks were created to standardize and compare the work of different investigators of this problem. The main objective of this work is to investigate how to combine data from diverse intrusion detection systems in order to improve the detection rate and reduce the false-alarm rate. From the statistical analysis we have proved the acceptability of FNNM in performance enhancement of intrusion de¬tection systems. The future improvements in individual IDSs can also be easily incorporated in this technique in order to obtain better detection ca¬pabilities. Experimental evaluation shows that the proposed methods have the capability of detecting a significant percentage of rare and new attacks.

6.References

[1] Wikipedia,The Free Encyclopedia, “Intrusion Detection System”, http://en.wikipedia.org/wiki/Anomaly-based_intrusion_detection_system, Accessed August 15, 2014.

[2] Wikipedia, The Free Encyclopedia, “Intrusion Detection System”, http://en.wikipedia.org/wiki/Misuse_detection Accessed August 15, 2014.

[3] Fatai Adesina Anifowose and Safiriyu Ibiyemi: EludioraApplication of Artificial Intelligence in Network Intrusion Detection, a Succinct Review, World Applied Programming, Vol (2), No (3), . Pp 158-166, March 2012

[4] Ravi Ranjan and G. Sahoo : A New Clustering Approach For Anomaly Intrusion Detection, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.4, No.2, March 2014

[5] K. M. Faraoun and A.Boukelif, Neural Networks Learning Improvement using the K-Means Clustering Algorithm to Detect Network Intrusions, International Journal of Computational Intelligence, pp161-168, 2007

[6] Mrutyunjaya Panda & Manas Ranjan Patra, Some Clustering Algorithms to Enhance the Performance of the Network Intrusion Detection System, Journal of Theoretical and Applied Information Technology, pp710-716, 2008

[7] Mostaque Md. Morshedur Hassan: Current Studies on Intrusion Detection System, Genetic Algorithm and Fuzzy Logic, International Journal of Distributed and Parallel Systems (IJDPS) Vol.4, No.2, March 2013

[8] Dragan Simic and svetlana simic, Advances in Intelligent and Soft Computing, Volume 95/2011, pp: 717-726,2011

[9] H. Gunes Kayacik, A. Nur Zincir-Heywood, and Malcolm I. Heywood, “On the capability of an SOM based intrusion detection system,” in Proceedings of the International Joint Conference on Neural Networks, IEEE, IEEE, vol. 3, July 2003, pp. 1808–1813.

[10] Hervé Debar and Andreas Wespi, “Aggregation and Correlation of Intrusion-Detection Alerts”, Recent advances in intrusion detection, Vol. 2212, 2001, pp. 85-10.

[11] J. McHugh, “Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA IDS evaluations as performed by Lincoln Laboratory”, ACM Transactions on Information and System Security, vol.3, Nov. 2000, No.4.

[12] J. E. Dickerson, Jukka Juslin, Ourania Koukousoula, Julie A. Dickerson, “Fuzzy Intrusion Detection”, Proc. Of 20th NAFIPS International Conference, Jul. 2001.

[13] H.S. Javitz, A. Valdes, “The NIDES Statistical Component Description and Justification,” Technical report, SRI International, Menlo Park, CA, March 1994.