Indian Journal of Science and Technology, *Vol 7(9), 1437–1443, September 2014*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Multi-Tier Framework using Sugeno Fuzzy Inference System with Swarm Intelligence Techniques for Intrusion Detection

S. Revathi* and A. Malathi

PG and Research, Department of Computer Science, Government Arts College, Coimbatore-18, India; revathisujendran86@gmail.com, malathi.arunachalam@yahoo.com

Abstract

An intrusion detection has a key role in network security that classifies the system activities as normal or suspicious (Anomaly). An intrusion detection system must consistently detect malicious activities in a network and must perform efficiently to manage with the large amount of network traffic. The main objective of this paper is to analysis two issues such as accuracy and efficiency of the system by a novel method of incorporating swarm intelligence with data mining algorithm for feature reduction. The accuracy of the system then can be achieved by several soft computing techniques as Sugeno fuzzy inference system and simplified swarm optimization. The high efficiency can be achieved by Multi-tier approach. The proposed system uses Multi-tier-sugeno fuzzy inference system for fuzzy rule generation that effectively identifying the intrusion activities within a network, Finally, to obtain best result Simplified swarm optimization algorithm are used to optimize the structure of the fuzzy decision engine. The experimentation and evaluation of the proposed method were performed on NSL KDD intrusion detection dataset that shows best accuracy and efficiency than other methods and can easily detect whether the network data are normal or under attack.

Keywords: Intrusion detection, Multi-Tier approach, NSL-KDD, Random Forest, Simplified Swarm Optimization, Sugeno Fuzzy inference system

1. Introduction

The pervasive use of computers and internet has enhanced the assets of many people's life, but it also exposed to security threats both externally and internally. The security of the system is compromised when an intrusion takes place[1]. The IDS with high Detection Rate (DR), and low False Alarm Rate (FAR) is a challenging task to be faced by researcher[2,3]. Additionally soft computing is also an innovative approach used to construct an intelligent system has also been applied to intrusion detection. There are several soft computing paradigms in which fuzzy logic plays a vital role in detecting intrusion[4,5]. Most fuzzy systems make use of human expert knowledge to create their fuzzy rule base and henceforth lack variation, therefore, building fuzzy systems with learning and adaptation capabilities has recently received much attention[6]. Two main aspects in data mining are data classification and feature selection. However, existing traditional data classification and feature selection techniques used in data management are no longer enough to detect intrusive data.

The proposed system has Multi-tier approach for IDS. A new concept of hybrid swarm optimization technique with random forest algorithm is used to extract features form audit data that used to classify network activities. The preprocessed data obtained is then moved to fuzzy inference system for rule generation or as a decision making engine to detect the network activities as normal or intrusive. Finally simplified swarm optimizations are used to optimize the proposed structure of fuzzy engine. The comparison of various IDS work initially based on DARPA dataset[11] developed by Lincoln Laboratory and

*Author for correspondence

Air force Research Laboratory and later the knowledge Discovery and Data mining has generated a TCP dump data that provides a network connection records for both training and testing named as KDD Cup 99 dataset[8]. The statistical issues that degrades the performance of KDD dataset lead to a replacement of NSL-KDD dataset which contains only selected records is used for our experimentation.

The rest of the paper is planned as follows: In Section 2, it discuss the related work of various researcher briefly. In Section 3, a detailed description of NSL KDD dataset with various attacks. Section 4 explains Fuzzy inference system and its approaches and section 5 describe the proposed concept with optimization techniques for our experimentation. In Section 6 the experimental analysis and result discussion for the proposed system is detail. Finally in section 7 draws some concluding remarks and future work of this research area.

2. Related Work

There are several relevant background information exist on IDS research which have been used in modelling survival. It analysis the concept, and the analytical methodologies used for finding intrusion. The analysis of KDD dataset and its inherent problem leads to new version of NSL KDD dataset that are mentioned in[7,12]. It is very difficult to signify existing original networks, but still it can be applied as an effective benchmark data set for researchers to compare different intrusion detection methods. In[7] they have conducted a statistical analysis on this data set and found two important issues which highly affect the performance of evaluated system, and results in very poor evaluation of anomaly detection approaches. To solve these issues, they proposed a new dataset, NSL-KDD, which consists of only selected records form the complete KDD dataset and does not suffer from any of the mentioned shortcomings.

Initially data mining has a major part in detection intrusion based on classification and machine learning techniques proposed by Lee et al.[9], then Association rules and Frequent Episodes algorithms have been used to develop correlations between features and to learn pattern classification for the audit records, respectively. Agarwal and Joshi proposed a framework for learning a rule-based model (PN rule) to make classifier models on a dataset that has widely different class distributions in training data[10]. Recently biological inspired approaches have been extensively originated in network intrusion pattern detection. The field of "swarm intelligence" has attracted an increasingly number of researchers since the proposal of Particle Swarm Optimization (PSO) algorithm and also of the Ant Colony Optimization (ACO) Algorithm. To overcome the drawback of partial swarm optimization Yeh et al.[14] proposed new method by combining Simplified swarm optimization with weighted exchange local search method for intrusion detection. Similarly Fuzzy rule-based classifiers, decision trees, SVM, linear genetic programming have been used in by Abraham and Jain[13] to show the importance of soft computing paradigm for modeling intrusion detection systems.

None of those previous reviews have listed or compared feature reduction based hybrid simplified swarm optimization with random forest algorithm and the proposed model of using Multi-tier approach based on sugeno fuzzy inference system with swarm optimization for intrusion detection based on NSL-KDD dataset.

3. Dataset Description

The NSL KDD dataset has asset of 41 attributes derived from each connection and the status of the each connection record label as normal or specific attack type. These features mainly based on continues, discrete and symbolic which falls under four categories of attack as Denial of Service, Probe, User to Root, Remote to Local, Normal. To solve statistical issues, that degrade the performance of KDD cup 99 dataset leads to evaluation of new data set as, NSL-KDD[7] is, which consists of only selected records of the complete KDD data set. The main advantage of NSL KDD dataset are[16]:

1. No redundant records in the train set.
2. No duplicate record in the test set.
3. The number of selected records from each difficult level group is inversely proportional to the percentage of records in the original KDD data set. Table 1 shows the major attacks in both training and testing dataset[7].

The training dataset is made up of 21 different attacks out of the 37 present in the test dataset. The known attack types are those present in the training dataset while the novel attacks are the additional attacks in the test dataset. The number of records in training and testing dataset is listed in Table 2.

4. Fuzzy Inference System

Fuzzy sets were introduced by Zadeh in 1965 to represent and manipulate data and information in which there are various alternative uncertainties. Among several combinations of methodologies in soft computing, the one that has the most important concept is fuzzy logic and neurocomputing, which leads to neuro-fuzzy systems. An effective method developed by Jang for this purpose is called ANFIS (Adaptive Neuro-Fuzzy Inference System). Fuzzy Inference System (FIS) is a computer pattern based on fuzzy set theory, fuzzy if-then-rules and fuzzy reasoning. It is a process of mapping given input to an output using fuzzy sets theory. Figure 1 shows an example of the fuzzy inference system with five functional blocks. The steps performed in the FIS are as follows:

- The fuzzification process transforms each crisp input variable into a membership grade based on the membership functions defined.
- The inference engine used for applying the appropriate fuzzy rule in order to obtain the fuzzy set to be accrued in the output variable.
- The defuzzifier transforms the fuzzy output into a crisp output by applying a specific defuzzification method.

Three well known types of fuzzy inference system that have been widely used in various application are Mamdani, Sugeno and Tsukamota Fuzzy inference system. The main difference between the three FIS is it lays consequents of their fuzzy rule but the aggregation and defuzzification procedures may differ.

Table 1. Attacks in Testing Dataset

Attacks in Dataset	Attack Type (37)
DOS	Back, Land, Neptune, Pod, Smurf, Teardrop, Mailbomb, Processtable, Udpstorm, Apache2, Worm
Probe	Satan, IPsweep, Nmap, Portsweep, Mscan, Saint
R2L	Guess_password, Ftp_write, Imap, Phf, Multihop, Warezmaster, Xlock, Xsnoop, Snmpguess, Snmpgetattack, Httptunnel, Sendmail, Named
U2R	Buffer_overflow, Loadmodule, Rootkit, Perl, Sqlattack, Xterm, Ps

Table 2. Dataset

Attacks	Training dataset	Testing Dataset
Normal	67343	9711
DOS	45927	45927
Probe	11656	2421
U2R	52	200
R2L	995	2756

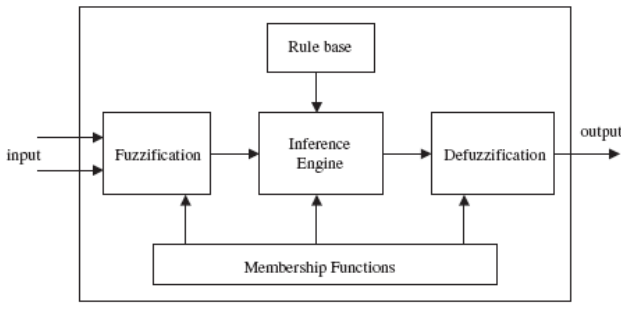

Figure 1. Fuzzy Inference System.

4.1 Mamdani Fuzzy Model

The Mamdani Fuzzy model is the first method proposed to map an input to output space. It was proposed by Ebrahim in 1975 based on Lotfi Zadeh's 1973 concept on fuzzy algorithms for complex systems and decision processes[17]. It based on consequents as fuzzy sets, and the final crisp output of the Mamdani method is based on defuzzification of the overall fuzzy output using various types of defuzzification methods, it mainly based on Centroid Of Area (COA). To completely specify the operation of a Mamdani fuzzy inference system, the function to be assigned for each of the following operators as AND operator (usually T-norm), OR operator (usually T-conorm), Implication operator, Aggregate operator and Defuzzification. The main drawback of Mamdani model is its lower computational efficiency than sugeno and time increases for aggregation and defuzzufication process.

4.2 Sugeno Fuzzy Inference System

The Sugeno fuzzy model was proposed by Takagi-Sugeno-Kang[18] is an effort to develop a systematic approach to generate fuzzy rule form given input-output dataset. The method is similar to Mamdani concept but it various in output membership function as either linear or constant. The sugeno fuzzy model has been written as

$$If\ x\ is\ A\ and\ Y\ is\ B\ then\ Z = f(x,y)$$

Where A and B are input fuzzy sets in antecedent and usually z = f(x, y) is a zero- or first-order polynomial function in the consequent. If the function produce a constant output that it belong to Zero order sugeno model or otherwise first order model. Fuzzy reasoning procedure for the first order Sugeno fuzzy model is shown in Figure 2, here, defuzzification procedure in the Mamdani fuzzy model is replaced by the operation of weighted average in order to avoid the time consuming procedure of the former[19].

The main Advantage of Sugeno Fuzzy Model are[19]:

- Sugeno is a more compact and computationally efficient representation than a Mamdani system.
- It works well with optimization and linear techniques (e.g., PID control).
- It is well suited to mathematical analysis

5. Proposed System

The proposed architecture for intrusion detection consist of three modules, Initially in the preprocessing phase the proposed SSO-RF algorithm completely reduce the attributes which clearly improves the detection rate, then in Multi-Tier approach where a number of security checks are performed in sequence. The main aim is to reduce computational time and to eliminate communication overhead between each tier. Every Multi-Tier framework is trained and tested separately using Fuzzy inference system to make a final decision for recognition. FIS implements nonlinear mapping based on multi-tier approach which specifies the input as normal or intrusive. Finally in order to attain best result simplified swarm technique is used to optimize the structure. Figure 3 depicts block diagram for proposed system.

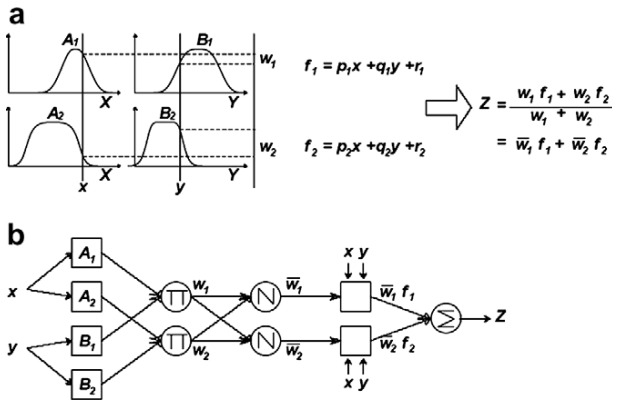

Figure 2. (a) The Sugeno fuzzy model reasoning; (b) equivalent ANFIS structure.

5.1 Feature Reduction Module

The NSL-KDD dataset has 41 attributes out of which some may be irrelevant or unwanted attribute for a specific attack, to reduce the attribute based on attack significance, this paper proposed a new concept of pre-processing approach that filters data effectively. The new proposed model namely Simplified Swarm Optimization (SSO)[15] is incorporated with random forest algorithm. Simplified swarm optimization is used for optimization of the dataset and random forest algorithm used for splitting of the dataset to reduce important features. The hybrid SSO-RF reduce the attribute to 11 which increase the accuracy and efficiency of the proposed system for intrusion detection.

This approach is used to solve classification problem and reduce dimensionality of dataset. Random Forest built a random sampling feature set for each node in a tree which reduces the correlation between the trees that improves the efficiency. The main advantage of using random forest is, it overcomes the problem of over fitting and the training data are less sensitive to outlier detection, finally it's easy to handle high dimensional, continuous, categorical and binary data[16]. The proposed SSO-RF method filters raw data and reduce irrelevant and dimensionality problem for both discrete and continuous variables in dataset. This approach is expressively different from other research work which had combined only data mining and PSO. The proposed system produces high accuracy and produce near optimal solution for pre-processing phase.

5.2 Multi- Tier Sugeno Fuzzy Decision Module

In many situations, there is an imbalance between efficiency and accuracy of the system and there can be

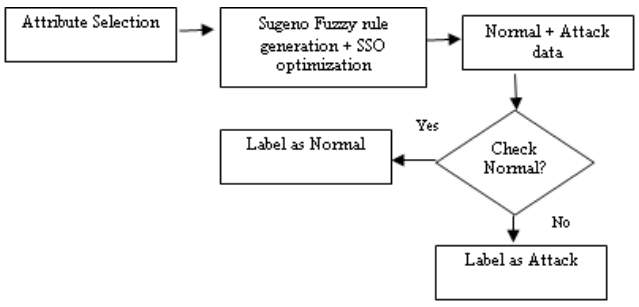

Figure 3. Block diagram for single Multi-Tier system.

various ways to improve system performance. Various methods are used to increase system efficiency. To balance this trade-off, we use the sugeno fuzzy inference system to generate various fuzzy rules that are integrated with multi-tier approach for various approach to improve more accurate and overall system performance. The algorithm for integrating Fuzzy inference system with Multi-Tier approach is given below:

Algorithm: Training Phase

Step 1. Select Multi-Tier, Say n, for the complete system.

Step 2. Perform Feature Selection for individual attacks in the Tier.

Step 3. Train a separate model with sugeno fuzzy inference system with fuzzy rule generation for individual attack in a Tier using the feature selection.

Step 4. Plug the trained module serially so that only normal is passed to next layer.

Algorithm: Testing Phase

Step 5. Test instance are performed from step 6 to step 9.

Step 6. Test the data are label either as attack or normal.

Step 7. If the data is identified as attack with its tier name then goto step 5 else move the sequence to next layer.

Step 8. If all the Tiers are not tested then move to step 7 else to step 9.

Step 9. To obtain the best result optimize the test data using simplified swarm optimization to the structure of fuzzy decision making and label the attack.

5.3 Simplified Swarm Optimization (SSO) Module

SSO is a simplified version of Partial swarm optimization and can be used to find the global minimum of nonlinear functions. Initially, the number of swarm population size, the number of maximum generation, and three parameters are determined. In every generation, the particle's position value in each dimension will be kept or be updated by its pbest value or by the gbest value or be replaced by new random value according to the procedure depicted in equation 1[14].

$$x_{id}^{t}=\begin{cases} x_{id}^{t-1} & if \quad rand() \in [0, c_w) \\ p_{id}^{t-1} & if \quad rand() \in [c_w, c_p) \\ g_{id}^{t-1} & if \quad rand() \in [c_p, c_g) \\ x & if \quad rand() \in [c_g, 1) \end{cases} \qquad (1)$$

Where i = 1; 2; m, where m is the swarm population. Xi position of particles. Cw, Cp and Cg are three predetermined positive constants with Cw < Cp < Cg. Pi = (pi1; pi2; : : : ; piD) denotes personal best and Gi = (gi1; gi2; : : : ; giD) denoted global best solution. The x represent new particle with random value between 0 and 1. The SSO algorithm is used to optimize membership function of fuzzy decision making module. The fitness function has been evaluated for each individual based on detection rate and false alarm rate to increase accuracy and efficiency of the proposed system.

6. Experimental Result and Analysis

The experimentation has been analyzed based on NSL-KDD dataset, a simplified version of KDD cup 99. The dataset consist of 125,973 training data and 22544 testing data are used to evaluate classifier. In the proposed method each rules are extracted only when statistical significant level occurs frequently. The test significant is carried out based on rules evaluated in training level. The result has been compared with other machine learning techniques and it's that the proposed multi-tier approach is higher in both accuracy and efficiency.

The experiment has been performed using MATLAB and WEKA tool for our integrating approach. We divide the train and test data into five different groups as Normal, Dos, probe, U2R and R2l. The attacks are experimented individually with normal data. The fitness function is used to calculate detection rate and false alarm rate. The DR is computed based on ratio between numbers of correctly detected attack with total number of attacks. The FAR is computed based on number of normal records incorrectly classified as attack with total number of attacks.

From the Table 3, it is clear that the proposed system significantly performs better than other existing machine learning approaches such as navie bayes and decision tree algorithm. The experimentation has also been carried to check individual attacks types for our proposed Multi-tier sugeno model and the overall performance has been analyzed and Table 4 shows the performance of the system.

As revealed by Table 4, the proposed system shows high detection and low false alarm using multi-tier sugeno fuzzy inference system. The training time and testing time has also been evaluated for individual attacks.

Table 3. Performance of proposed system

Approach	DR	FAR	Train Time (sec)	Test Time (sec)
Multi-tier Sugeno Model	98.44	1.02	5.78	2.11
Multi-Tier Decision Tree	89.23	2.14	2.43	4.23
Multi-Tier Navie Bayes	86.12	2.98	1.90	3.87

Table 4. Individual attacks performance for Multi-Tier sugeno model

Attacks	DR	FAR	Train Time (sec)	Test Time (sec)
Normal	99.2	0.8	6.9	4.2
DOS	98.7	0.9	8.76	9.23
Probe	97.6	1.2	6.88	5.76
U2R	97.8	1.3	8.90	9.87
R2l	98.9	0.9	5.89	2.90

7. Conclusion

In this paper, an evolutionary soft computing approach has been introduced for detecting intrusion based on training and testing NSL-KDD dataset. The Sugeno model is capable of generating fuzzy rules without human effort. A fuzzy decision making engine was used to make system more powerful in detecting attacks. This paper also proposed Simplified Swarm Optimization method to optimize the Membership Function of fuzzy decision module to produce better accuracy and increase efficiency of the system to detect attacks in computer network. Filters such as Kalman, Extended Kalman and other noise signal filtering can be used to filter attack data. The implementation can also be proposed to real time network attack data to produce very high performance in detecting attacks.

8. References

1. Heady R, Luger G, Maccabe A, Servilla M. The architecture of a network level intrusion detection system. Technical report. Computer Science Department, University of New Mexico: 1990 Aug.
2. Tiawan D, Abdullah AH, Dris MY. Characterizing Network Intrusion Prevention System. Int J Comput Appl. 2011 Jan; 14(1):11–18.
3. Axelsson S. 2000. Intrusion detection systems: A survey and taxonomy. Department of Computer Engineering, Chalmers University; Report No.:99–15.
4. Zadeh LA. Role of soft computing and fuzzy logic in the conception, design and development of information/ intelligent systems. Lect Notes Comput Sci. 1998; 695:1–9.
5. Gomez J, Dasgupta D. Evolving fuzzy classifiers for intrusion detection. Proceeding of 2002 IEEE Workshop on Information Assurance; 2001 Jun: West Point, NY, USA: United States Military Academy; p. 68–75.
6. Gao M, Zhou MC. Fuzzy intrusion detection based on fuzzy reasoning Petri nets. Proceeding of the IEEE International Conference on Systems; 2003 Oct 5–8; Washington, DC, USA: Man and Cybernetics; 1272–7.
7. Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A Detailed Analysis of the KDD CUP 99 Data Set. Proceeding of the IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009); 2009 Jul 8–10; Ottawa, ON. IEEE; 2009. p. 1–6.
8. KDD Cup 1999 Intrusion detection dataset. Available from: http://kdd.ics.uci.edu/ databases/kddcup99/kddcup99.html
9. Lee W, Stolfo SJ, Mok K. A data mining framework for building intrusion detection models. Proceedings of IEEE Symposium on Security and Privacy; 1999 May 9–12; Oakland, CA, USA. IEEE; 1999. p.120–32.
10. Agrawal R, Imielinski T, Swami A. Mining Association Rules between Sets of Items in Large Databases. Proc ACM SIGMOD. 1993; 22(2):207–16.
11. McHugh J. Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory. ACM Trans Inform Syst Secur. 2000; 3(4):262–94.
12. Nsl-kdd data set for network-based intrusion detection systems. 2009 Mar. Available from: http://nsl.cs.unb.ca/KDD/ NSL-KDD.html
13. Abraham A, Jain R. (2005) Soft computing models for network intrusion detection systems. Studies in Computational Intelligence. 2005 Aug 22; 4:191–207.
14. Bae C, Yeh W-C, Wahid N, Chung YY, Liu A. A new simplified swarm optimization (sso) using exchange local search scheme. International Journal of Innovative Computing, Information and Control. 2012 Jun; 8(6):4391–406.
15. Yeh WC, Chang WW, Chung YY. A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. Expert Syst Appl. 2009 May; 36(4):8204–11.
16. Revathi S, Malathi A. Optimization of KDD Cup 99 Dataset for Intrusion Detection Using Hybrid Swarm Intelligence with Random Forest Classifier. Int J Adv Res Comput Sci Software Eng. 2013 Jul; 3(7):1382–7.

17. Mamdani EH, Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. Int J Man Mach Stud. 1975; 7(1):1–13.
18. Takagi T, Sugeno M. Fuzzy identification of systems and its applications to modeling and control. IEEE Transaction on Systems, Man, and Cybernetics. 1985 Jan/Feb; 15:116–32.
19. Jang JSR. ANFIS: Adaptive-Network-based Fuzzy Inference Systems. IEEE Transactions on Systems, Man, and Cybernetics. 1993; 23:(3):665–85.