Intrusion Detection Based On Fuzzy Logic Approach Using Simplified Swarm Optimization

S. Revathi [#1], Dr. A. Malathi[*2]
[1]*Ph.D. Research Scholar,* [2]*Assistant Professor*
PG and Research, Department of Computer Science
Government Arts College
Coimbatore-18

Abstract--The intrusion is becoming more essential for effective defense against attacks that are constantly changing in magnitude and complexity. Mainly intrusion detection relies on the extensive knowledge of security experts. The paper proposed a new detection mechanism as Fuzzy Intrusion Detection Engine (FIDE) that uses fuzzy logic to access network data. FIDE uses fuzzy analyzer engine to evaluate inputs and generate alerts for security administrators. The FIDE act as a fuzzy classifier, whose knowledge base is act as fuzzy "if-then" rule. This paper describes the components of FIDE architecture, and explains the benefit of fuzzy rule that improve fuzzy sets. Finally, in order to obtain the best result Simplified Swarm Optimization is used to optimize the structure of FIDE. The simulation of the proposed system is trained and tested with actual real time network data. The FIDE IDS can detect a wide range of common attack types. The proposed system shows high accuracy in identifying attacks.

Keywords: Fuzzy Logic, intrusion Detection, Simplified Swarm Optimization, FIDE.

I. INTRODUCTION

The development of internet in computer system leads to intrusion detection a more remarkable attention [1]. The internet access turns computer system more susceptible to attack due to its network connectivity. The ID is used to monitor system and network from malicious activities. The two main intrusion detection models are Misuse and Anomaly detection approaches. Misuse detection relies on matching known patterns against databases for past attacks. Although it can be quite effective at identifying only known attacks and their variants, it's unable to detect new security attacks and require an ongoing threat update to remain effective. Anomaly detection describes normal behavior of the user and analyze the behavioral changes based on statistical measures to compare current activity with historical knowledge [2].

Current intrusion detection techniques, mainly based on discovering abnormal system events in computer networks and distributed communication systems. Due to the un-certainty nature of intrusions, fuzzy sets, play a vital role in identifying dangerous events and reducing false alarms level [3]. This paper discusses new mechanism of Fuzzy based network intrusion detection system to identify abnormal activities in real time computer system. The Fuzzy Intrusion Detection Engine (FIDE) uses a fuzzy system to identify malicious network activities. The system combines network traffic with fuzzy rules to determine general network attack. The training dataset is classified into subsets based on various attacks generated. The important attributes are identified and the actual dataset has been reduced, then fuzzy rules are generated based on input fuzzification to obtain if-then rule with the consequent part that represent whether data is normal or attack. The rest of the paper is organized as: Section II describes some related work based on a fuzzy concept for intrusion detection. Section III explains proposed system architecture. The implementation of network intrusion detection based on fuzzy logic and result analysis are described in section IV and section V. Section VI draws some conclusion and future work

II. RELATED WORK

Intrusion detection has developed as a substantial field of research, because it is not possible to set up a system with no vulnerabilities. One main conflict in intrusion detection is that we have to find out the hidden attacks from a large quantity of routine communication activities [4]. Several machine learning (ML) algorithms, for example Neural Network [5], Support Vector Machine [6], Fuzzy Logic [7], Data Mining [8] and Genetic Algorithm [9], and more have been extensively employed to detect intrusion activities both known and unknown from the large quantity of complex and dynamic datasets. Generating rules are dynamic for IDSs to differentiate standard behaviors from abnormal behavior by examining the dataset which is a list of tasks created by the operating system that are registered into a file in historical sorted order [12]. Various researches with data mining as the main component has been carried to find out newly encountered intrusions [10]. The analysis of data to determine relationships and discover concealed patterns of data which otherwise would go unobserved is known as data mining. Many researchers have cast-off data mining to focus into the subject of database intrusion detection in databases [11].

III. PROPOSED SYSTEM ARCHITECTURE

The Fuzzy Intrusion Detection Engine (FIDE) relies on fuzzy rule. FIDE consist of three components shown in Figure1. The network server data (NSD) are an immoral network data recorder or sniffer. It reads raw network packets on transmission and store them on disk. The next component, the Data Preprocessor (DP) which review and arrange the raw data packets into selected categories. In this paper a new proposed concept as Simplified Swarm Optimization with Random Forest algorithm is used to mine the raw data and reduce the attributes. These preprocessed data are fuzzified as input to the fuzzy threat analyzer (FTA) which then produce fuzzy alters based on condition to a degree shown in Figure1.

The main goal of FIDE is

- To demonstrate how fuzzy systems used in intrusion detection method.
- To identify which data sources are used as best inputs to the fuzzy intrusion detection system.
- To show fuzzy methods as best for network input data.
- To show how the system can be scaled to distributed intrusion detection involving multiple hosts and/or networks.

The First step in any intrusion detection methodology is to identify data feed and sources of information for IDS. The real time network server data collect data packet that crosses the through wire and store them on disk. No processing, filtering, or reducing of the data is performed by the NSD collector, Such Network data are collected from Linkware Technologies as live dataset. Next the raw data may contain various attribute which may be irrelevant for intrusion detection, so data preprocessing is proposed to mine the raw data. In step three the preprocessed data are given as a fuzzy input for fuzzy rule generation. Finally, fuzzy alert is generated.

IV. NETWORK INTRUSION DETECTION USING FUZZY RULE GENERATION

Recently in many research areas Fuzzy ideas and fuzzy logic are so often utilized to generate various rules. In this research work Fuzzy concept is used for intrusion detection using data mining and optimization techniques. The system makes use of various effective rules to identify attacks, which is obtained from by mining data effectively. The fuzzy rule concept is mainly used to provide better classification accuracy and to detect attacks. The different steps involved in detection are as follows.

4.1 Data Preprocessing
4.2 Classification of training data
4.3 Automatic fuzzy rule generation
4.4 Using SSO for optimization

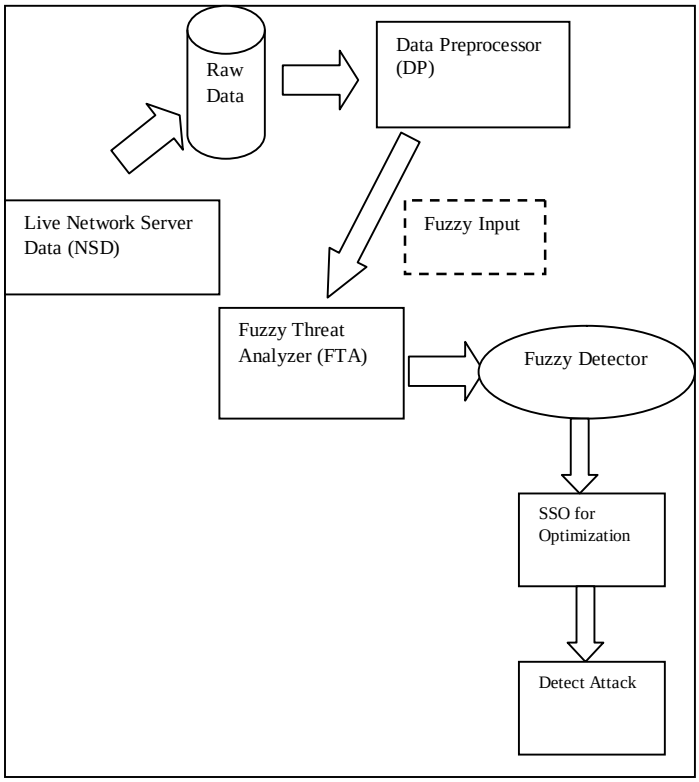

Figure1: Fuzzy Intrusion detection Engine (FIDE)

A. *Data Preprocessing Unit*

The traditional pre-processing algorithms are not adapted to the situations when dataset size is large. This may result in the false recommendations. In this paper, a new proposed hybrid technique are used, SSO is a simplified version of PSO which can be used to find the global minimum of nonlinear functions [14]. This approach is used to solve the classification problem and reduce the dimensionality of the dataset. The random forest classifier is used to split the dataset and to identify the most important attribute to detect intrusion [13]. On preprocessing the proposed method choose most suitable attribute for identifying the classification whether the record is normal or attack. All the 41 attributes are not so effective in detecting intrusion. The proposed method reduced the attribute to 11 that are taken for automatic fuzzy rule generation to train the system [17].

B. *Classification of Training Data*

The first module of the proposed system is about classifying the input data that contains various attacks. The dataset used is a real time network server data which involves different attacks. Initially we categories these

server data into five different attacks as DOS, Probe, User to Root, Remote to Local and Normal with 41 attributes that have real valued attributes. The proposed system reduced attribute to 11 which is taken for training phase. The dataset (D) is divided into five subsets of the class label as D= {D_i; where i= 1≤i≤5}. The class label describes several attacks, which comes under four major attacks (Denial of Service (DOS), Remote to Local (R2L), User to Root (U2R) and Probe) along with normal data. The five datasets are then used for generating a better set of fuzzy rules automatically so that the fuzzy system can learn the rules effectively.

C. Automatic Fuzzy Rule Generation

In general fuzzy rules are generated manually by experts which may be time consuming and difficult to analyze intrusion behavior. But in our case, generating fuzzy rule is more difficult due to real time network data which is huge in size and also having more attributes, so we used automatic fuzzy rule generation to find a better set of rules. The definite rule is generated based on Lagrange's Interpolation using successive approximation method.

Generating Fuzzy Rule: In proposed system the fuzzy rules are automatically generated using Lagrange's Interpolation with successive approximation method. The fuzzy rules are generated from the definite rules, where the antecedent part of the rule is a numerical variable and consequent part is a class label related to attack name or normal. But, the fuzzy rule is linguistic variable. So, to make the fuzzy rules from the definite rules, we fuzzified the numerical variable of the definite rules and THEN part of the fuzzy rule is same as the consequent part of the definite rules. For example, "IF attribute1 is Very High (VH), then the data is attack and "IF attribute1 is Very Low (VL), then the data is normal".

The process of a fuzzy system based on

1. Fuzzification: Convert classical data or crisp data into fuzzy data or Membership Functions (MFs)
2. Rule Evaluation: Each fuzzy rule is determined based on degree of membership of crisp input value in the fuzzy set of antecedent.
3. Defuzzification: It transposes the fuzzy outputs into crisp values.

D. Simplified Swarm Optimization Module

The Simplified Swarm optimization is used to solve linear optimization problems based on swarm population size, maximum generation and three predetermined constant [14,15]. The particle position is updated based on pbest or gbest value or by random number depicted in eq 1. It selects individual based on optimal solution from the current particle. The algorithm stops till stopping criteria met. In the proposed system, each individual particle has coded parameters of the MFs of the input fuzzy set of the fuzzy decision engine. The fitness function evaluates the fitness value for each individual. Essentially, the fitness function is the function that should be optimized based on detection rate and false alarm rate for evaluating intrusion detection.

Fuzzy Fitness Function: To reduce the computational complexity of optimization technique based on fitness function, this paper used a new concept of adaptive fuzzy fitness granule queue based on insufficient similarity, fitness is not interpolated or estimated; rather the uncertainty in the similarity among real solutions is exploited. In this method an adaptive queue of solutions (fuzzy granules) with an exactly computed fitness function is maintained.

$$\left(X_j^i\right) = \begin{cases} f(X_d) & if \max_{d\in\{1,2,\ldots,L\}} \{\bar{\mu}_{j,d}\} > \theta^i \\ f(X_j^i) & computed\ by\ fitness\ function, otherwise \end{cases}$$

$$where\ d = \underset{d\in\{1,2,\ldots,L\}}{index\ max} \{\bar{\mu}_{j,d}\}, \quad (2)$$

$$X_j^i = \{X_{j,1}^1, X_{j,2}^1, \ldots X_{j,m}^1, \ldots X_{j,n}^1\}$$

If new individual is similar to known fuzzy granule, it used as a crude estimate otherwise it added to individual fuzzy granule. The main aim of fitness granule is to reduce the number of exact fitness function by creating queue [16]. When a new solution is introduced to this queue, granules compete by a measure of similarity to win the new solution and thereby to prolong their lives in the queue. In turn, the new individual simply assumes fitness of the winning (most similar) individual in this queue. If none of the granules are sufficiently similar to the new individual, i.e. below a certain threshold value, then the new individual is added to the queue after its fitness is exactly evaluated by known fitness function. Finally, granules that cannot win new individuals are gradually eliminated in order to avoid a continuously enlarging queue. Based on these fitness function detection rate and false alarm rate are calculated

V. SIMULATION RESULT ANALYSIS

To evaluate IDS, the system uses 30 % of training dataset that contains both normal and four attack data, which are given to the proposed system initially for identifying the suitable attributes. Those selected attribute of the rule generation process is given in table 1. These attributes are given to fuzzy rule learning strategy, so that the system generates both definite and indefinite rules and finally, fuzzy rules are generated from the definite rules.

In the testing phase, complete dataset is evaluated to classify the input data as a normal or attack. The obtained result is then used to compute overall accuracy of the proposed system, which are normally used to estimate the rare class prediction. The detection rate and false positive rate are calculated based on

$$DR = \frac{Total number of correctly classified attacks}{Total number of instance} * 100$$

$$FAR = \frac{Total number of misclassified instance}{Total number of instance} * 100$$

Table 1: Selected attribute for rule generation

Selected Features (SSO_RF)

1. Protocol Type- Discrete
2. Service Name- Discrete
3. Flag status- Discrete
4. Size of Source in Bytes- Continuous
5. Size of Destination in Bytes- Continuous
6. Emergency Flag- Discrete
7. Connection Status- Discrete
8. Number of Outbound commands- Continuous
9. Accessed Files Count- Continuous
10. Host Login Status- Discrete
11. Guest Login Status- Discrete

The evaluation metrics are figured for both training and testing dataset and the obtained results for all attacks and normal data are given in table 2, which is the overall classification performance of the proposed system on live network dataset. By analyzing the result, the overall performance of the proposed system is significantly improved and it achieves more than 75% accuracy for all types of attacks.

Table 2: Performance of Proposed System

Model	Class Name	Accuracy (DR)	FAR	Train Time(sec) 30%	Test Time (sec) Full Dataset
Fuzzy	Normal	74.42%	2.54%	71.1	202.4
	DOS	74.62%	2.54%	66.1	201.2
	Probe	73.62%	2.58%	71.1	202.4
	U2R	74.82%	2.52%	67.1	201.5
	R2L	74.84%	2.49%	67.3	201.5

VI. CONCLUSION

This paper proposed a new concept of using fuzzy rule generation with a soft computing approach to detect intrusion behavior based on the live network dataset. An automatic fuzzy rule generation makes the system more accurate in detecting attacks. Finally to improve detection, this paper proposed to use simplified swarm optimization to optimize fuzzy decision making engine. The proposed fitness granule queue makes the IDS more optimized. The accuracy and false alarm rate be around 75 % and 2.54%, still the training and testing time for fuzzy rule generation are needed to be reduced, which can be focused as future work, further improved by using hybrid filters such as kalman, unscented kalman filter etc., which is our future work to be focused to make intrusion detection system more effective in computer networks.

References

1. H. Debar, M. Dacier, A. Wespi, Towards a taxonomy of intrusion detection systems, Computer Networks 31 (1999) 805–822.
2. G. Macia Fernandez and E. Vazquez, “Anomaly-based network intrusion detection: Techniques, systems and challenges”, Computers & Security, Vol. 28, No. 1-2, pp. 18-28, February-March 2009.
3. J. Luo, “Integrating fuzzy logic with data mining methods for intrusion detection,” Master’s thesis, Dept. Comput. Sci., Mississippi State Univ., Starkville, MS, 1999.
4. S. Axelsson, 2000. Intrusion detection systems: a survey and taxonomy, Department of Computer Engineering, Chalmers University, Report No. 99-15.
5. S.-J. Han and S.-B. Cho, “Evolutionary neural networks for anomaly detection based on the behavior of a program,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 36, no. 3, pp. 559–570, Jun. 2006.
6. Q. Tran, H. Duan, and X. Li, “One-class support vector machine for anomaly network traffic detection,” presented at the 2nd Netw. Res. Workshop 18th APAN, Cairns, Australia, Jul. 2004.
7. J. Gomez, D. Dasgupta, “Evolving fuzzy classifiers for intrusion detection”, in: Proceeding of 2002 IEEE Workshop on Information Assurance, United States Military Academy, West Point, NY, USA, 2001, pp. 68–75.
8. D. Barbara, J. Couto, S. Jajodia, L. Popyack, and N. Wu, “ADAM: Detecting intrusions by data mining,” in Proc. 2nd Annu. IEEE Workshop Inf. Assur. Secur., New York, Jun. 2001, pp. 11–16
9. D. E. Goldberg, “Genetic Algorithm in Search, Optimization and Machine Learning. Reading”, MA: Addison-Wesley, 1989.
10. A. A. Freitas, “Data Mining and Knowledge Discovery with Evolutionary Algorithms”. New York: Springer-Verlag, 2002.
11. W. Lee and S. Stolfo, “Data mining approaches for intrusion detection,” in Proc. 7th USENIX Secur. Symp., San Antonio, TX, Jan. 1998, pp. 79–83.
12. W. Lee, S. Stolfo, and K. Mok, “A Data Mining Framework for Building Intrusion Detection Model”, In Proceedings of the IEEE Symposium on Security and Privacy, Oakland, CA, pp. 120-132, 1999.
13. Arif Jamal Malik, Waseem Shahzad, Farrukh Aslam Khan. “Binary PSO and Random Forests Algorithms for PROBE attacks Detection in a network”. In Proceedings of IEEE Congress on Evolutionary Computation, 662-668, (2011).
14. Yao Liu, Yuk Ying Chung, Wei-Chang Yeh: “Simplified Swarm Optimization with Sorted Local Search for golf data classification”. IEEE Congress on Evolutionary Computation (2012): 1-8.
15. S.Revathi, A.Malathi, “Data Preprocessing for Intrusion Detection System using Swarm Intelligence Techniques”, International Journal of Computer Applications (0975 – 8887) Volume 75– No.6, August 2013.
16. M. Davarynejad, M.-R. Akbarzadeh-T, IEEE Senior Member, N. Pariz, “A Novel General Framework for Evolutionary Optimization: Adaptive Fuzzy Fitness Granulation”.
17. Safaa Zaman, Mohammed El-Abed, Fakhri Karray, “Features Selection Approaches for Intrusion Detection Systems based on Evolution Algorithms”, ICUIMC(IMCOM)’13, Kota Kinabalu, Malaysia, ACM 978-1-4503-1958-4…. January 17–19, 2013.