DOI 10.29042/2019-5031-5035

Helix Vol. 9 (3): 5031- 5035

A Hybrid Database Intrusion Detection Algorithm Using Swarm Intelligence and Radial Basis Function Network

*[1]Anitarani Brahma, [2]Suvasini Panigrahi, [3]Jayasmita Mahapatra
[1,2]Veer Surendra Sai University of Technology, Burla, Odisha
[3]iNurture Education Solution Pvt Ltd
Email: brahmaanita00@gmail.com, suvasini26@gmail.com,jayasmitamahapatra2010@gmail.com

Received: 27th December 2018, Accepted: 13th February 2019, Published: 30th June 2019

Abstract
Recently, all over the globe, intrusion detection especially in Database gaining lots of attentions from researchers as it serves a security support system to existing security mechanism. This paper presents a hybrid intrusion detection algorithm in Database System by making use of swarm intelligence and radial basis function network. Both these techniques are combined to produce better detection results. The performance of the proposed approach is investigated with synthetic dataset and is compared with other machine learning technique and found to be significantly better

Keywords
Swarm Intelligence, Radial Basis Function Network, Particle Swarm Intelligence, Database Intrusion Detection

Introduction

Due to advancement of Information Technology in the field of e-commerce, online banking , online marketing, the existing database expose to the outside world which contains many sensitive information that should not be disclosed. Hence, intensive care is needed to the existing security mechanism and Database Intrusion Detection System (DIDS) serves additional protection to the Database. The main objective of DIDS is to discriminate between normal and malicious transactions made to the database. Many statistical mechanism already been applied by researchers to develop an accurate DIDS model [1,2,3,4,5,6]. Most of the researches on DIDS fall in the category of anomaly based detection as they detect intrusive attack if they deviate from the normal behavioral access pattern and they do not rely on signature type attack that is in case of misuse detection category. It is common and general that, anomaly based intrusion detection have to cope with large volume and high dimensional data gathered from multiple sources and has to deal with real dynamic environment. This is the reason, the computational Intelligence come into play in the field of intrusion detection. To achieve real time database intrusion detection, researchers investigated so many methods to improve the accuracy. But this may not be sufficient to secure the dynamic real time database.

Intrusion Detection is considered as classification problem by researchers using soft computing techniques such as Support Vector Machines, Artificial Neuro Fuzzy Inference System, Fuzzy Logic, Neural Network, Genetic Algorithm, Decision Tree [7]. A probabilistic database intrusion technique is proposed by S.Cho by using Self Organizing Map and fuzzy logic [8]. To reduce false alarm, Panigrahi et al. utilized fuzzy logic to develop database intrusion detection behavioral model [9]. . Recently, Swarm Intelligence (SI) techniques have been adapted by research community in the field of intrusion detection specially in Network [10,11]. Swarm Intelligence is a measure of collective behavior of ant, fish, bird or other societies of animal to implement algorithm. The advantage of Swarm Intelligence can be a candidate solution to DIDS. In order to improve the accuracy of Intrusion Detection System (IDS), multiple popular soft computing techniques can be hybridized. Due to good generalization capabilities and sparse representation of solutions, Radial Basis Function Network and Particle Swarm Optimization techniques are combined to develop a novel DIDS model.

The objectives of the paper are as follows, first, to build an accurate DIDS model which classifies between the normal transactions and malicious transactions. Second, modeling of normal behavior of database users accurately so as to reduce the false positives. Third, utilizing the concept of swarm intelligence in the field of DIDS to see the performance deviation of DIDS compared to other soft computing techniques like Support Vector machines, ANFIS.

The paper outline is as follows, proposed framework with little highlight of methods used is discussed in section 2, followed by the experimental setup and evaluation in section 3. Conclusion and future work are outlined in section 4.

Proposed Approach

This investigation illustrates the anomaly intrusion detection where in the training phase; a number of records are fed to the detection engine. A classification algorithm is applied in this component to discriminate the incoming behavior as normal or abnormal. So, intrusion detection can be sensed and reduced to classification

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

problem. In this context, researchers always crave for easy-to-implement, fast, real-time, efficient methods for DIDS to challenge the real world fraud cases. The unique characteristics of Swarm Intelligence (SI) and Radial Basis Function Network (RBFN) make it ideal for this purpose. Following sections describe about the methods used and the proposed algorithm for DIDS.

Swarm Intelligence: For solving complex problem, human always follows Nature. Biology inspired approaches become most popular in the fields of research like engineering, computer science, economics, medicine and social sciences. Swarm Intelligence is one of the biology inspired techniques. In the global optimization framework, G. Beni and J. wang first introduced SI as a set of algorithm to control robotic swarm [12]. Ant Colony Optimization (ACO), Particle Swarm Intelligence (PSO) and Bee Colony Optimization (BCO) are the popular SI techniques. ACO which was proposed by Dorigo and colleagues[13] inspired by collective behavior of Ant that is ant-foraging behavior, brood-sorting behavior, cemetery formation behavior and cooperative transport behavior[14] for solving combinatorial optimization problem. PSO is a population based optimization technique introduced by J. Kennedy et al [15] inspired by the social behavior of bird flocking and fish schooling. D. Karabago proposed BCO algorithm by observing the reaching a nectar source of bee by following a mate who has already discovered the path to destination [16]. This investigation utilizes PSO to optimize the parameters used in RBFN based DIDS model.

Particle Swarm Optimization: Particle Swarm Optimization (PSO) algorithm is a fast optimization algorithm based on the group intelligence behavior of bird flocking. It has high convergence speed, global search capability and easy to implement strategy. This searching process starts with an initial population called Swarm which has their random solution called particle and each iteration, they move forward to the solution. Two fitness values are to be updated in each iterations of all particles: first, personal best (pbest) which is the best solutions or performance of itself and second, best position obtained by whole swarm called gbest. Calculation of velocities should limited to [-Vmax, Vmax] [17]. Updating of velocities and position are done as per Eq. (1) & (2).

$$Vid(t+1)=\omega*Vid(t)+c1r1(Pid(t)-Xid(t))+c2r2(Pgd(t)-Xid(t)) \quad (1)$$

$$Xid(t+1) = Xid(t) + Vid(t+1) \quad (2)$$

Radial Basis Function Network: RBFN is a3-layer artificial neural network proposed by Broomhead and Lowe has been used in many applications such as function approximation, pattern recognition, time series prediction and classification [18]. The input layer feeds the data to the hidden layer; neurons in the hidden layer are activated based on the distance between each input pattern with the centroid stores in the hidden layer. The centriods are also called RBF neurons. Different types of activation functions may be used in the hidden layer, but Gaussian function as equated in Eq.(1) is mostly used. The output layer sums the hidden neuron values and calculates the output as equated in Eq.(3).

$$\phi = \exp\frac{-(x-cj)^2}{2\sigma j^2} \quad (3)$$

Where X= input pattern (x1,x2,x3…..,xn), j=1,2,…n
Cj= Center of the jth hidden neuron j= width of the jth neuron
The output of the RBF network is calculated as in Eq. (4)

$$Y = \sum_{k=1}^{m} W_{jk}\phi \quad (4)$$

Where Wjk= Weight between jth hidden layer to kth output layer (k=1,2,….m)

To increase the accuracy of RBFN, fitness function measured that takes error between actual and real output should be reduced. Root Mean Square Error (RMSE), Sum Square Error (SSE), Normalized Root Mean Square Error (NRMSE) are the fitness function has been used by the researchers. We use MSE to increase the accuracy as described in Eq. (5)

$$MSE = \frac{\sum_{i=1}^{n}(T-Y)^2}{N} \quad (5)$$

The performance of RBFN is purely depends on the suitable number of neurons in the hidden layer and the weights between hidden and output layer.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

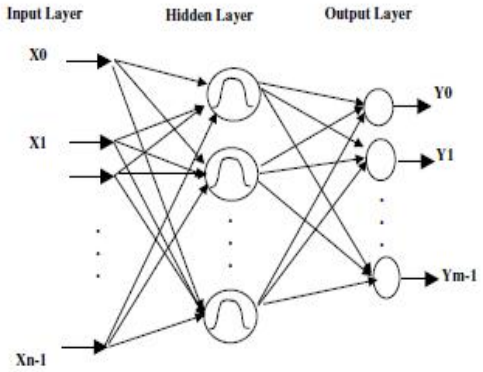

Figure 1: Data Flow Diagram of Radial Basis Function Neural Network

Other than any Artificial Neural Network, RBF has several advantages such as it can model any nonlinear function using a single hidden layer (which eliminates of calculation of number of hidden layer and nodes), the solution doesn't fall in local minima problem, less computation needed while calculation of weight metric between hidden and output layer. But in RBFN, some preprocessing required on the training and test data.

Hybrid_RBFN_PSO: Determination of best values of RBF centers can be accomplished through PSO which ultimately improves the performance of classification. To optimize radii and weight, we use K- nearest Neighbourhood and Singular Value Decomposition (SVD) algorithm

Algorithm:

- Randomly initialize center.
- Initialize Radii r using KNN
- Use SVD to initialize Weights w.
- Start optimizing centers of RBFs using PSO.
- Initialize particles position and velocity randomly
    - While not reach the maximum numbers of iteration do for each particle do Calculate fitness value (MSE between Real output of RBFN and Target)
        - If fitness value is better than best fitness value of pbest in particle, then Set current position as pbest

      End
    - Select gbest of the particle which the best fitness value among all particles in current iteration
        - Calculate particle velocity based on (1).
        - Update particle position (centers) based on (2).

End

End

- Take the gbest of particle as centers c of RBF
- Complete training RBFN using K-NN and SVD.
- Calculate the real output of RBFN
- Repeat

Proposed Approach:

Considering the advantages of the above hybrid classifier, we proposed to apply this hybrid neural network classifier to detect intrusions in database. To increase the performance of the DIDS model, the first task is preprocessing of input data. This input data are normalized by max-min normalization and 10-fold cross modulate to generate training and test dataset because of their good performance and simplicity. RBFN based DIDS model is trained using the feeded training data where the RBFN parameters are optimized using PSO by following the procedure above. By imparting training, RBFN now perform as anomaly based classifier which can classify the normal and intrusive transactions very efficiently. The advantages of this proposed model is its real time behavior which can adapt itself with dynamic scenario. Performance of proposed classifier for DIDS is tested through several performance measures.

Results and Discussion

The proposed anomaly based DIDS classifier is a hybrid combination of RBFN and PSO. RBFN is the primary classifier which is trained to classify normal and intrusive transactions. PSO is used to optimize its performance.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

Along with these two techniques KNN and SVD techniques are used for determination RBF parameter to increase the performance. As there is no real ready mate dataset found in the literature or any other source, the model is evaluated using a simulator generated dataset which can generate synthetic transactions of legitimate and intrusive transactions [9]. In order to demonstrate the performance of the proposed model, we have taken 80000 normal and attack records. This input data are preprocessed by following normalization and cross validation procedure. 9 fold data are used for training purpose where as 1 fold is used for testing the classifier. The input data have the following features of transaction:

<user-id, transaction-id, sequence of table, sequence of attribute, date, time, location, attribute_count based on their sensitivity level>

In the experiments, Detection rate (DR) and False Positive Rate (FPR) are the two performance measures to assess its performance. Especially, FPR is very essential to the performance of a DIDS. High number false alarms which are prohibited may generate than actual number of alarms through small difference of FPR. Ratio of correctly detected attacks with the total number of attacks is labeled as Detection Rate; whereas, the ratio between incorrectly classified normal transactions with total number of normal transactions is identified as False Positive Rate. For evaluation of classifier algorithm, another performance measure is taken in this investigation is Cost Per Example (CPE) as equated in (6).

$$CPE = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{m} CM(i,j) * C(i,j) \quad (6)$$

where CM= Confusion Matrix, C=Confusion matrix, N= Total number of instances

A confusion matrix is a square matrix where each entry represents the number of misclassified instances that originally belong to i incorrectly classifies as j. Each entry in C matrix represents the cost penalty for misclassifying an instance.

Parameter	Value
Number of particle	20
Number of iterations	500
C1	0.5
C2	2
Vmax	1

Table 1: Parameters for PSO Used in Hybrid-RBFN- PSO-DIDS

Method	DR (in %)	FPR (in %)	CPU Time in millisecond
Hybrid_RBFN_PSO_DDS	97.2	2.5	30

Table 2: Performance of the Proposed Approach

In the rest of this section, the performance of the proposed Hybrid_RBFN_PSO_DIDS is discussed. Table 1 listed the value of parameters used in PSO in this investigation. It is evident from the Table 2, that proposed model produces admirable detection rate, reduced false positive rate and also take less CPU time due to less number of hidden layer and the optimized value of RBF parameter. The proposed approach has 0.168 CPE. Results shown above clearly describes that the proposed approach has better performance while detecting intrusion in Database.

Conclusion

In this paper, two soft computing techniques for database intrusion detection were employed and the model was successfully demonstrated its usefulness on a synthetic dataset. RBF network was used as a classifier for intrusion detection in database. PSO is used in RBFN to improve its performance in clarity of decision making process. Experimental results showed that the proposed DIDS in detecting intrusions in Database. Our future work will focus on building accurate DIDS model by combining both misuse and anomaly based detection process by embedding several soft computing techniques

References

1. V. Lee, J. Stankovic, S. Son, “Intrusion Detection in Real time Databases via Time Signatures”. In: Proceedings of the 6th IEEE Real-Time Technology and Applications Symposium, RTAS, pp. 124–133 .2000.

2. A. Rakesh , S. Ramakrishnan, “Fast algorithms for mining association rules”. In: Proc. of 20th International Conference on Very Large Data Bases. Berlin: Morgan Kaufmann, pp.487-499,1994.

3. D. Barbara, R. Goel, S. Jajodia, “Mining Malicious Data Corruption with Hidden Markov Models”, In: Proceedings of the 16th Annual IFIP WG 11.3 Working Conference on Data and Application Security, July 2002, pp. 175–189 ,2002.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

4. Y. Zhong, X. Qin, " Database Intrusion Detection Based on User Query Frequent Itemsets Mining with Item Constraints", In: Proceeding of the 3rd international conference on information security, pp. 224–225 ,2004.
5. E. Bertino, E. Terzi, A. Kamra, A. Vakali, "Intrusion Detection in RBAC-Administered Databases", In: Proceedings of the 21st annual computer security applications conference (ACSAC), pp. 170–182 ,2005.
6. S. Panigrahi, S. Sural, A.K. Majumdar, "Two-stage database intrusion detection by combining multiple evidence and belief update", Inf Syst Front,2010.
7. S.X. Wu, W. Banzhaf, "The Use of Computational Intelligence in Intrusion detection systems: A Review", Applied Soft Computing, 10, pp. 1-35, 2010.
8. S. Cho, "Incorporating Soft Computing Techniques Into a Probabilistic Intrusion Detection System",IEEE Transactions on Systems, Man and Cybernetics-Part C, Application and Reviews, 32, pp.154-160, 2002.
9. S. Panigrahi, S. Sural, "Detection of Database Intrusion Using Two-Stage Fuzzy System", LNCS 5735,pp. 107-120, 2009.
10. C.Kolias, G. Kambourakis, M. Maragoudakis, " Swarm intelligence in Intrusion detection: A survey", Computers & Security, 30, pp.625-642, 2011.
11. M. Sailaja, R. Kiran Kumar, P. Sita Rama Murthy, P.K. Prasad, "A Novel Approach for Intrusion Detection Using swarm Intelligence", Proceedings of the InConINDIA, AISC 132, pp. 469-479,2012.
12. G. Beni, J. Wang , "Swarm intelligence in cellular robotic systems Proceedings of NATO Advanced Workshop on Robots and Biological Systems", 102, 703-712.1989.
13. M.Dorigo, V. Maniezzo, A. Colorni, " Positive feedback as a search strategy", Technical Report 91-016, Dipartimento di Elettronica, Politecnico di Milano, Italy,1991.
14. O.A. Mohamed Jafar, R. Sivakumar, " Antbased Clustering Algorithms: A Brief Survey". International Journal of Computer Theory and Engineering, 2(5) 1793-8201,2011.
15. J. Kennedy,R.C. Eberhart , "Particle swarm optimization", Proceedings of IEEE International Conference on Neural Networks, 1942–1948,1995.
16. D. Karaboga. An idea based on honey bee swarm for numerical optimization, Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department,2005.
17. D. Niu, Y. Lu, X. Xu, & B. Li , "Short-term power load point prediction based on the sharp degree and chaotic RBF neural network". Mathematical Problems in Engineering, 2014.
18. J.S.R. Jang , "Neuro-fuzzy and soft computing", PHI publication, 1st edition, Pearson education, New Delhi, 2004.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495