2581-4575

International Journal For Recent Developments in Science & Technology

Crossref

A Peer Reviewed Research Journal

# PRODUCTIVITY INTRUSION DETECTION SYSTEM USING GENETIC ALGORITHM

**[1]DR.P.AVINASH, [2]K.RAJASHEKHAR RAO**

[1]Professor, Dept of CSE, Sridevi Women's Engineering College, Hyderabad.

[2]Asst Prof, Dept of CSE, Sridevi Women's Engineering College, Hyderabad.

**ABSTRACT:** The crucial part of building lightweight IDS depends on preprocessing of network data, identifying important features and in the design of efficient learning algorithm that classify normal and anomalous patterns is identify the abnormal activities in the network traffic and security network resources and illegal penetrations by intruders. A new intelligent Conditional Random Field (CRF) based feature selection algorithm to change the number of features. In existing Layered Approach (LA) based algorithm is used to perform classification with these reduced features. This paper deals with Intrusion Detection System method to classification. KNN is applied to binary classifier for anomaly detection. We present an Intrusion Detection System (IDS) is applying genetic algorithm (GA) to efficiently detect many types of network intrusions. Parameters and evolution processes for GA are discussed in details and implemented. A hybrid misuse intrusion detection model is made to find attacks on system to improve the intrusion detection. Based on prior features, intrusion on the system is detected without any previous learning. The major advantages of this proposed system are reduction in detection time, increase in classification accuracy and reduction in false alarm rates.

**Index Terms:** Intrusion detection system, false alarms, LA, intelligent CRF, K-Nearest Neighbor (KNN), Neural Network (NN), Genetic Algorithm. Feature Selection

## 1. INTRODUCTION

Recent years, due to dramatic growth in networked computer resources and variety of network based applications is developed to provide services in many different areas [1]. The increase in the number of networked machines is lead to an increase in unauthorized activity not only from external attackers but also from internal attackers such as disgruntled employees and people abusing their privileges for personal gain [2]. In the current internet world intrusion detection [3] is one of the high priority and challenging tasks for network administrators and security professionals. Even in the presence

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

of more sophisticated security model the attackers come up with newer and more advanced penetration methods to defeat the installed security systems [4]. In KNN classification an object is classified by a majority vote of its neighbors. The object is consequently assigned to the class that is most common among its Kith Nearest Neighbor, in K is a positive integer that is typically small [5]. The prominence factor of network based intrusion detection system (NIDS) is that any single instance of time, NIDS can monitor multiple systems in a network in parallel. The network is helps in monitoring the information traversing through the network and detects any adversary to intrusion activities [6].

## 2. RELATED WORK

Present latest approaches to the process of detecting intrusions utilize some form of rule-based analysis. Rule-Based analysis relies on sets of predefined rules and provided by an administrator automatically created by the system. Expert system is most common form rule-based intrusion detection system [7]. We propose new Intelligent CRF based LA (LAICRF) model is developed by combining an Intelligent CRF based Feature Selection Algorithm (ICRFFSA) and LA based classification algorithm for effective intrusion detection. This model uses intelligent agents is capable of sensing the environment and perform actions based on the environmental conditions [8]. N. Suguna [9] proposed a method Genetic Algorithm (GA) and K-Nearest Neighbor (KNN) is combined to improve classification performance. Instead of considering all the training samples and taking K-neighbors the GA is employed to take K-neighbors straightway and the distance to classify the test samples is calculated [10]. The effort of using GAs for intrusion detection is referred back [11] applied the multiple agent technology and GP to detect network anomalies [12]. For two agents they used GP to determine anomalous network behaviors and each agent is monitor one parameter of the network audit data. The NSL-KDD data set that has been used for intrusion detection is refined data set of original KDD'99 data set KDD'99[5] data set was having the problem of duplicate or redundant data is removed in NSL-KDD [13] data set.

## 3. SYSTEM ARCHITETURE

The architecture of the proposed system for effective intrusion detection. It consists of four major components namely knowledge base, feature selection module that contains a feature selection agent, intrusion detection module is training agent and decision making agents. All these components is responsible for performing intrusion detection effectively [14]. These

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

positions could be referred to as genes. An evaluation function is used to calculate the goodness of each chromosome according to the desired solution this function is known as Fitness Function [15].

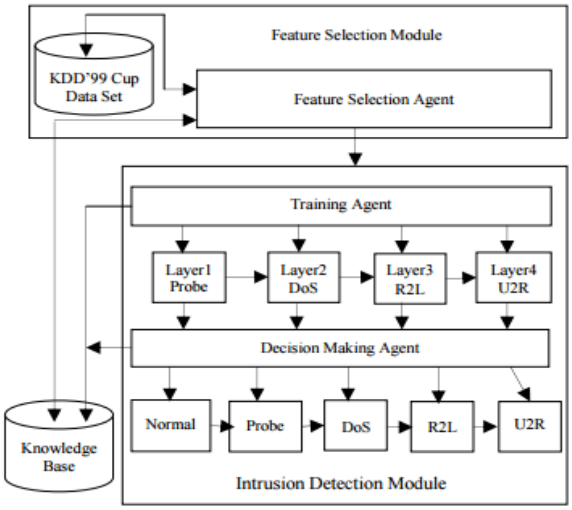


Figure 1. System Architecture

## 4. PROPOSED SYSTEM

The proposed system has three different phases. CRF is type of probabilistic system that is used to model the conditional distribution of random variables of any order. CRF is an unbiased and undirected graphical model that can be used to perform sequence labeling [16]. In our hybrid KNN_NN two layer models is use both the concept of K-Nearest Neighbor and Neural Network classifier. KNN is used for binary classification which classifies data into normal and abnormal classes. Then the abnormal classes of KNN are passed to Neural Network for classifying specific attack type [17]. We use GA for solving many problems three factors is impact on the effectiveness of the algorithm and also of the applications [18]. They are: i) the fitness function; ii) the representation of individuals; and iii) the GA parameters. The determination of these factors often depends on applications and implementation.

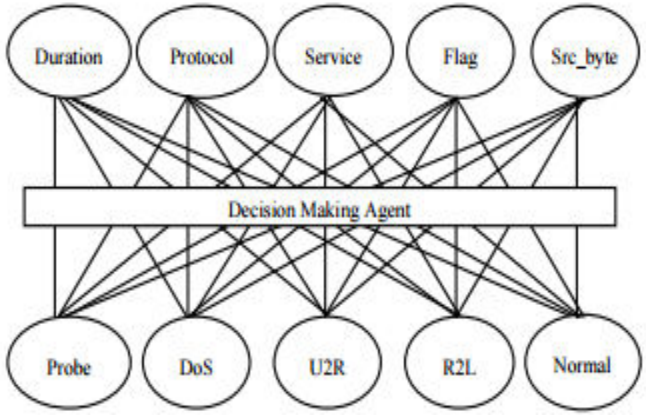


Figure 2. Graphical representation of CRF

### A. Neuron tree Algorithm

The predicted results produced by these neural networks are combined based on the voting algorithm. Then, the trained NN ensemble is employed to generate a new training set through replacing the desired class labels of the original training [19].

Begin

1. Get input file Ti for training

2. Read records from Ti

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

3. Train the network by specifying the number of input nodes, hidden nodes, output nodes, learning rate and momentum.

4. Initialize weights and bias to random values.

5. Calculate output for each node

6. Calculate Error rate (ER) = E (FP, FN) Therefore, Error rate = w ∗ FP + w ∗ FN

7. Output cell error=Logistic function derivative * Error rate

8. Hidden Cell error = Logistic function derivative* Sum of (output layer cell error *weight of output layer cell connection)

9. Net weight = Current Weight between hidden layer and output + (output cell error * hidden layer cell value *learning rate)

10. Net Bias value = Current bias Value + (learning rate *output cell error)

11. Training is completed.

12. Return trained neural network.

End.

The information gain measure used in enhanced algorithm is used to select the test attribute at each node in the tree. Such a measure is referred to as an attribute selection measure or a measure of the goodness of split

### B. CRF Based Feature Selection Algorithm

Feature selection is the process of selecting appropriate features from the underlying data set such as KDD'99 cup data set for building models [20]. In the CRF based feature selection algorithm, each feature is added to class of values depending on their dependency information. However, this dependency information is computed based on random values. Therefore, to improve the efficiency of feature selection, we propose an intelligent agent and CRF based feature selection.

**Algorithm: Intelligent CRF based feature Selection.**

**Input:** The set S of all features

**Output:** F, the set of optimal features

Begin

F= { }; // Initialize F to all null set.}

For i=1 to n do

Begin for j=1 to n do

Begin

f=random(S, CRF(s)) //Feature Selection

CV=CV+Cond.prob(fi )

D=DA(CV, Decision)

if decision=="yes" then F=F∪(fj )

Val=Check (CV >Threshold(Ai )) and

Constraints (i, j))

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal


If (val==true)

Display (Ai , j, Features(S));

Prevent (Ai , j);

Else Stop

End

End

End

### C. Algorithm: Suggested-KNN

We used feature selection to minimize memory space required for large dataset and to reduce learning time. Also, to produce much better result we have designed hybrid KNN_NN model instead of using separated KNN and Neural Network classification method.

**Input:** NSL-KDD for training and testing.

**Output:** Results of anomaly detection on NSL-KDD testing dataset using KNN.

1. Suppose there are training categories and the sum of the training samples is , after feature reduction, they become m-dimension feature vector.

2. Make sample to be the same feature vector of the form as all training samples.

3. Calculate the similarities between all training samples and Taking the sample as an example; the similarity is calculated using Euclidean distance.

4. Choose k samples which are larger from similarities of and treat them as a KNN collection of Then calculate the probability of belong to each category respectively [21]..

5. Judge sample to be the category which has the largest probability

6. End

## 5. GENETIC ALGORITHM

GA uses an evolution and natural selection that uses a chromosome-like data structure and evolve the chromosomes using selection, recombination and mutation operators [22]. The process usually begins with randomly generated population of chromosomes, which represent all possible solution of a problem that are considered candidate solutions. From each chromosome different positions are encoded as bits, characters or numbers. These positions could be referred to as genes.

**Algorithm: Predict data type (using GA)**

**Input:** Network audit data (for testing), Recalculated set of chromosomes

**Output:** Type of data.

1. Initialize the population

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal


2. Cross over Rate = 0.15, Mutation Rate = 0.35

3. While number of generation is not reached

4. For each chromosome in the population

5. For each recalculated chromosome

6. Find fitness

7. End for

8. Assign optimal fitness as the fitness of that chromosome

9. End for

10. Remove some chromosomes with worse fitness

11. Apply crossover to the selected pair of chromosomes of the population

12. Apply mutation to each chromosome of the population

13. End while

## 1. HYBRID ALGORITHM

In the proposed hybrid algorithm first we will use the NSL-KDD data set; this data set will be processed in WEKA tool over machine learning algorithms.

1. Select training and test data (NSL-KDD data)

2. Generating rules over features of network

3. Building frame and making new rules by analyzing the generated rules.

4. Apply KDD data set and test the model.

5. Check evidence of attack, Calculate False positive and Negative

From processing different rules are produced by analyzing generated rules, network features of data set and results rules will be built for the network features.

## 6. EXPERIMENTATION RESULTS

All the models represented in table 1. works on large number of parameters to detect the intrusions but the proposed hybrid model contains only 29 features to detect the attacks which is far less than the available models and total number of features in the data set. The new hybrid model where no training is required for intrusion detection and other bars are representing the comparison with the existing models, in mostly all the models previous training is required and model are successful only on already known attacks

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

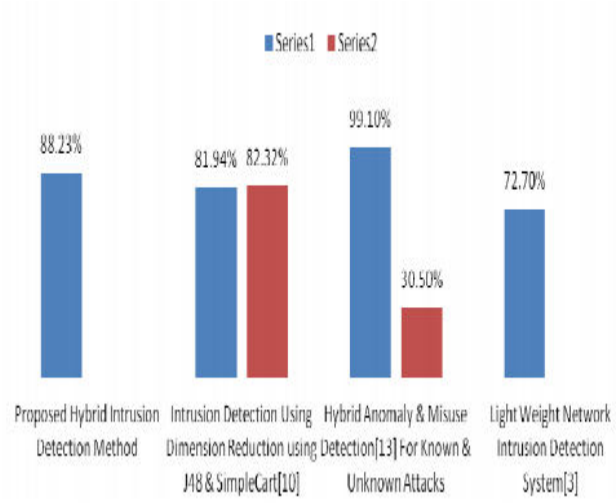


Figure 3. Comparative Analysis In Terms of Detection Accuracy

## 7. CONCLUSION

Model for intrusion detection shows that by analyzing the basic behavior of network data; based on prior features, and by hybrid model the machine learning algorithms intrusion detection can be improved. We can use a better equation or heuristic in this detection process we believe the detection rate and process will improve a great extent, especially false positive rate will surely be much lower. we proposed a LAICRF model which is developed by combining an ICRFFSA and LA based classification algorithm for effective intrusion detection.. Then classification method is performed on 25 selected features. Selection has been done by Rough Set Theory and Information Gain separately. The result shows that there is a significant decrease in memory space and also decrease in learning time of the algorithm as well as increase in the accuracy. Future extension to this work can be the use of temporal models to perform effective temporal reasoning based on time take to neuron tree model is employed as the classification engine which will improve detection rate.

## 8. REFERENCES

[1] S. Pontarelli, G. Bianchi, S. Teofili, Traffic-aware design of a highspeed fpga network intrusion detection system, Computers, IEEE Transactions on 62 (11) (2013) 2322–2334.

[2] Richard Power. 1999 CSI/FBI computer crime and security survey. Computer Security Journal,Volume XV (2), 1999

[3] Bace R. and Mell P., “Intrusion Detection Systems,” Computer Security Division, Information Technology Laboratory, Nat’l Inst. of Standards and Technology, 2001.

[4] Chimphlee W., Abdullah A., Sap M., Chimphlee S., and Srinoy S., “A Rough-Fuzzy Hybrid Algorithm for Computer Intrusion Detection,” the International Arab Journal of Information Technology, vol. 4, no. 3, pp. 247-254, 2007.

[5] Araujo J. D, Abdelouahab Z. Virtualization in Intrusion Detection System: A Study on Different Approaches for Cloud Computing

Environments. International Journal of Computer Science and Network Security (IJCSNS), Vol.12, No.11, pp.9-16(2012)

[6] Sinclair, C., Pierce, L., & Matzner, S. (1999). An application of machine learning to network intrusion detection. In Computer Security Applications Conference, 1999.(ACSAC'99) Proceedings. 15th Annual (pp. 371-377). IEEE.

[7] Kapil Kumar Gupta, Baikunth Nath, Ramamohanarao Kotagiri,‖ Layered Approach Using Conditional Random Fields for Intrusion Detection‖ IEEE Transactions On Dependable And Secure Computing, Vol. 7, No. 1, January-March 2010

[8] Datti R, Verma B. Feature Reduction for Intrusion Detection Using Linear Discriminant Analysis. International Journal on Computer Science and Engineering (IJCSE), Vol.2, No.4, pp.1072-1078(2010).

[9] Lakhina S, Joseph S, Verma B. Feature Reduction using Principal Component Analysis for Effective Anomaly–Based Intrusion Detection

[10] Suguna N, Thanushkodi K. An Improved k-Nearest Neighbor Classification Using Genetic Algorithm. International Journal of Computer Science Issues (IJCSI), Vol.7, Issue 4, No.2, pp.18-21(2010).

[11] M. Crosbie, E. Spafford, "Applying Genetic Programming to Intrusion Detection", Proceedings of the AAAI Fall Symposium, 1995.

[12] T. Xia, G. Qu, S. Hariri, M. Yousif, "An Efficient Network Intrusion Detection Method Based on Information Theory and Genetic Algorithm",

[13] Kim, G., Lee, S., & Kim, S. (2014). A Novel Hybrid Intrusion Detection Method Integrating Anomaly Detection with Misuse Detection. Expert Systems with Applications, 41(4), 1690-1700.

[14] M. Saniee Abadeh, J. Habibi, C. Lucas, "Intrusion detection using a fuzzy genetics-based learning algorithm", Journal of Network and Computer Applications

[15] Tao Peng, C. Leckie, Kotagiri Ramamohanarao, "Information sharing for distributed intrusion detection systems", Journal of Network and Computer Applications,

[16] Prema L. and Kannan A., "An Active Rule Approach for Network Intrusion Detection with Enhanced C4.5 Algorithm," the Journal of Communications, Network and System Sciences, vol. 1, no. 4, pp. 314-321, 2008.

[17] Ibrahim L. M, Basheer D. T, Mahmod M. S. A Comparison Study For Intrusion Database

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal


(KDD99, NSL-KDD) Based On Self Organization Map (SOM)

[18] R. H. Gong, M. Zulkernine, P. Abolmaesumi, "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", 2005

[19] K.K. Gupta, B. Nath, R. Kotagiri, and A. Kazi, ―Attacking Confidentiality: An Agent Based Approach,‖ Proc. IEEE Int'l Conf. Intelligence and Security Informatics (ISI'06), vol. 3975, pp. 285-296, 2006.

[20] Mccallum A. and Sutton C., "An Introduction to Conditional Random Fields for Relational Learning," Introduction to Statistical Relational Learning, vol. 4, no. 4, pp. 267-373, 2006

[21] Han J, Kamber M, Pei J. Data Mining Concepts and Techniques. Elsevier Book, 3rd edition (2012).

[22] J. P. Planquart, "Application of Neural Networks to Intrusion Detection", SANS Institute Reading Room.