International Journal of Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.06,Issue.29
September-2017,
Pages:5720-5724

www.ijsetr.com

Construction of Recognition Methodology using Genetic Algorithm

G. KRISHNA VENI[1], N V S K VIJAYALAKSHMI K[2], T. SATYA NAGAMANI[3]

[1]Assistant Professor, Dept of IT, Sir C. R. Reddy College of Engineering, Eluru, AP, India.
[2]Assistant Professor, Dept of IT, Sir C. R. Reddy College of Engineering, Eluru, AP, India.
[3]Assistant Professor, Dept of IT, Sir C. R. Reddy College of Engineering, Eluru, AP, India.

Abstract: System and network technology provide great convenience for rapid development. Network infrastructure devices help greatly in managing various spines of any enterprise like bank accounts, online transaction, transportation, social insurance, protection, correspondence and computer networks systems. Though many intrusion detection systems have been proposed in the past, the existing network intrusion detections have limitations in terms of detection time and accuracy. To overcome these drawbacks, we propose a new intrusion detection system in this paper by developing a new intelligent Conditional Random Field (CRF) based feature selection algorithm to optimize the number of features. In addition, an existing Layered Approach (LA) based algorithm is used to perform classification with these reduced features. One of the transformative methodologies for valuable choice is hereditary calculation Genetic Algorithm which is utilized as a hunt strategy while selecting highlights from full NSL(Nevada Seismological Learning) KDD information set. Support Vector Machines (SVM) has become one of the popular ML algorithm used for intrusion detection due to their good generalization nature and the ability to overcome the curse of dimensionality. As quoted by different researchers number of dimensions still affects the performance of SVM-based IDS. The results show that classification is done with high classification rate and low misclassification rate with the reduced feature subsets.

Keywords: Intrusion, Detection, Inspections, Terminology, Signature, Security, Alerts, Feature Selection, Genetic Algorithm (GA), Intrusion Detection System (IDS), Knowledge Discovery Data Mining Dataset (KDD), Support Vector Machine.

I. INTRODUCTION

The network intrusion detection system (NIDS) to detect possible intrusions as a malicious movement, hacking or policy misuse, a movement of a virus and alert proper uniquely to recognize [1]. The packets of data traveling over a network looking for suspicious activity monitored and analyzed by an NIDS [2]. The adaptation of ternary content (TCAM) addressable memories requires a deep packet inspection which involves exploring each byte of the packet's payload [3]. The signature based systems work with known signature patterns that are predefined by the administrator. They focus mainly on the known traffic data and from that the systems analyze the unwanted traffic [4]. In this scenario, the intrusion detection system should have a complete database that consists of all possible attacks, that helps to find the attacks easily. On the other hand, the anomaly based systems are the ones which are having collection of normal data [5]. To defeat this issue one can utilize less number of components than initially to decrease the measurements of the dataset [6].Selecting features is one of the procedures of dimensionality lessening, in which just restricted components are chosen utilizing an algorithmic methodology disregarding unimportant elements. By methodologies Intrusion Detection System (IDS) is the framework which recognizes offensive action on the system [7].These attacks range from relatively benign ping sweeps to sophisticated techniques exploiting security vulnerabilities [8]. To defend various cyber attacks and computer viruses, lots of computer security techniques have been studied in last decade, which include cryptography, firewalls and intrusion detection system (IDS) Among these techniques, intrusion detection has been more promising for defending complex and dynamic intrusion behaviors [9].

II. RELATED WORK

A network intrusion detection systems use anomalies based systems. Methods analyzed the popular movement of the network with normal network traffic and located any analytical anomalies [10]. Only the exceptional movement is identified as a potential exploitation. The flow of network traffic is monitored and compared by an IDS with a conventional baseline. The Denial-Of-Service [DOS] is a cyber-attack where the perpetrator attempts to make a computer network device unavailable to its expected users by momentarily or generally disrupting settings of a host connected to the Internet [11]. We propose an Intelligent CRF based LA (LAICRF) model which is developed by combining an Intelligent CRF based Feature Selection Algorithm (ICRFFSA) and LA based classification algorithm for effective intrusion detection [12]. This model uses intelligent agents which are capable of sensing the environment and perform actions based on the

environmental conditions[13]. Correlation based FS determines the value of a subset of features utilizing heuristic strategies. Feature which is very connected with a class is considered as great and chose. In every subset quality are chosen by considering the level of excess in the middle of them and prescient capacity of every individual component [14]. Thus, there is a need to characterize a proper relationship measure which can list most critical and profoundly viable highlights [15].The predictive accuracy wasclassified as a means to evaluate the "goodness" of a feature set distance measures to compute the relevance of a set of features. These approaches suffer from many drawbacks the first major drawback is that feeding the classifier with arbitrary features may lead to biased results and hence we cannot rely on the classifier's predictive accuracy as a measure to select feature [16].

III. SYSTEM ARCHITECTURE

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 which has the training agent and decision making agents [17]. All these components are responsible for performing intrusion detection effectively. Another highly performing method is Artificial Neural Networks (ANN) which can model both linear and non-linear patterns. The resulting model can generate a probability estimate of whether given data matches the characteristics that it has been trained to recognize [18]. Feature selection is a process that selects a subset of features from input data, such as network traffic to reduce overheads of data processing and to improve the accuracy of attack detection; moreover, dimensionality reduction also decreases the computational load of models [19].

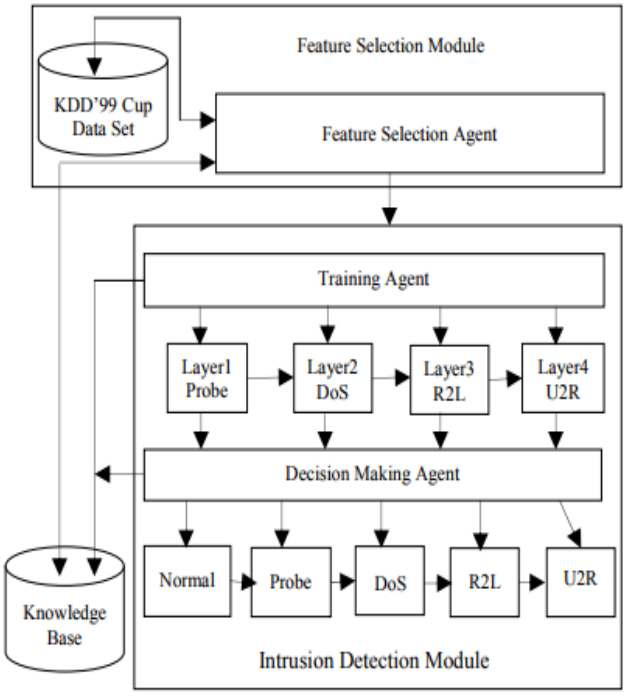

Figure 1. System Architecture.

IV. PROPOSED WORK

CRF (Conditional Random Field)are a type of probabilistic system [20] that is used to model the conditional distribution of random variables of any order. A CRF is an unbiased and undirected graphical model that can be used to perform sequence labeling. Our proposed approach for selecting necessary features for every layer and explain how some features were chosen over others [21]. We observe that the attack groups are different in their impact and it becomes necessary to treat them differently. We have selected features for each layer based upon the type of attacks that the layer will detect based on training [22]. This research presents a complete framework to select the best set of NSL-KDD dataset features that efficiently characterize normal traffic and distinguish it from abnormal traffic using Support vector machine. This research uses hybrid approach for feature selection that combines the filter and wrapper models.

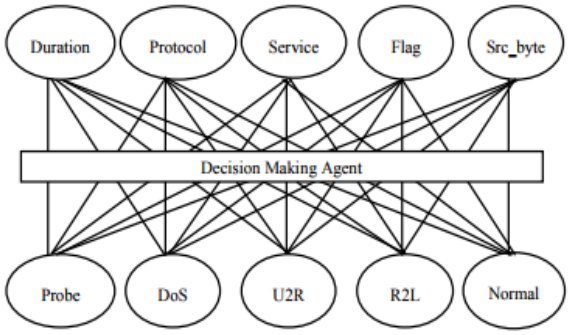

Figure 2. Graphical representation of CRF.

V. GENETIC ALGORITHMS(GA)

Genetic algorithms (GA) are an adaptive heuristic search method based on the idea of natural selection [23]. They are inspired by Darwin's theory of evolution survival of the fittest which is one of the randomized search techniques. The algorithm begins with a set of individuals called as population. Individual chromosome consists of a set of genes that could be bits, numbers or characters. Higher the fitness value more is the chances of an individual being selected. Crossover and mutation is responsible for producing new population. Crossover accelerates the search early in the evolution of the population while mutation is responsible for restoring the lost information to population global movement in the search space [24]. The process is iteratively repeated several times until stopping criteria are met or optimal solution is reached.

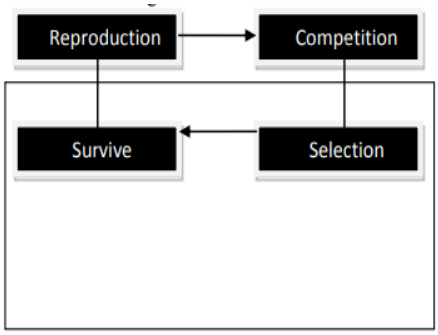

Figure3.Genetic Algorithms

A. Best Features Subset Selection

The k-means classifier is used to compute the detection rate for each subset of features. Initially, the set of features S contains only the top ranked feature [25]. A new feature is added to the list S based on the rank which it is assigned by the IGR measure one by one as long as the accuracy of selected subset is non-decreasing on features rank in both approaches. When the accuracy drops as an indication of model over fitting the algorithm is stopped.

Best Feature Selection Algorithm:

Input:

F – Full feature set

IGR: Information Gain Ratio Measure

C: K-means classifier

T: Gained Accuracy Threshold

For each feature f compute IGR(f)

Output:

S – Best feature subset

Algorithm

Initialize: S={}, ac=0

Repeat

- Assign acp= ac
- Evaluate f=getNext(F)
- Calculate S=SU{f}
- Calculate F=F-{f}
- Evaluate ac = accuracy(C,S)
- Continue the above steps until (ac-acp) < T Or ac <acp

This reduced feature set is then given as input to SVM for IDS model building. As the reduction model uses filter and wrapper approach, the efficiency and Intrusion detection capability of SVM model remained unchanged even with reduced feature set.

B. Machine Learning Algorithm

One of the rule-based methods which is commonly used by early IDS is the Expert System (ES) the knowledge of human experts is encoded into a set of rules. This allows more effective knowledge management than that of a human expert in terms of reproducibility, consistency and completeness in identifying activities that match the defined characteristics of misuse and attacks ES suffers from low flexibility and robustness.Decision Trees are one of the most commonly used supervised learning algorithms in IDS, high detection accuracy and fast adaptation. Another highly performing method is Artificial Neural Networks (ANN) which can model both linear and non-linear patterns [26]. For unsupervised intrusion detection, data clustering methods is applied. These methods involve computing a distance between numeric features and therefore they cannot easily deal with symbolic attributes resulting in inaccuracy. Support Vector Machines (SVMs) are also a good candidate for intrusion detection systems is provide real-time detection capability, deal with large dimensionality of data. SVMs plot the training vectors in high dimensional feature space through nonlinear mapping and labeling each vector by its class. The data is then classified by determining a set of support vectors, which are members of the set of training inputs that outline a hyper plane in the feature space [27].

VI. COMPARISION RESULTS

We used the benchmark KDD'99 cup intrusion data set for carrying out the experiments. This data set is drawn from DARPA intrusion detection system. The data set contains about five million connection records as training data, and two million connection records as the test data. Each record is unique in the data set continuous and nominal features plus one class label. We compare the proposed work with the existing classification methods namely LACRF and decision tree. From the experiments carried out it has been observed that LAICRF performs better than the other two techniques in terms of detection accuracy. The main reason for this improvement is that the proposed LAICRF not only insist that the observation features to be independent but also uses of intelligent agents to make effective decisions.

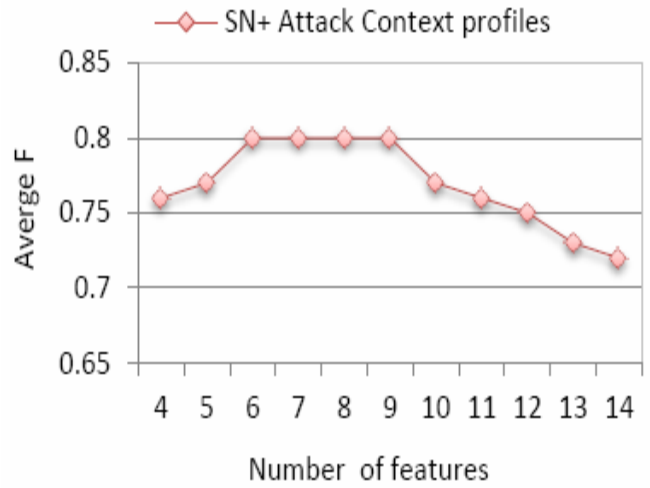

Figure4.The classification rate, misclassification rate

VII. CONCLUSIONS AND FUTURE WORK

The present research providesa system detection of intrusions and how many ways we can implement it and which will be useful for those beginners who are interested in the field of development of network intrusion detection system. According to previous approach, numerical convergence guideline based imaginative methodology utilizing Genetic Algorithm (GA) for feature selection. Feature selection is done utilizing distinctive feature selection (FS) techniques like CFS, IG and CAE and their impact on the execution of two ordinarily utilized

classifiers.We proposed an improved feature selection algorithm for network intrusion detection that performs data reduction by selecting important subset of attributes. The performance of our proposed approach on the KDD datasets achieved Stability and robustness for DOS attack class. The future work refers to multiclass classification problem, large amount of unlabeled data for training and incremental learning problems.his work can be the use of temporal models to perform effective temporal reasoning based on time. A series of new products referred to as system intrusion prevention, which not only detects an attack but also to take action appropriate on certain attacks has been recently delivered.

VIII. REFERENCES

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

[2] A. V. Aho and M. J. Corasick, "Efficient String Matching an aid to bibliographic search," Programing Techniques, vol. 18, no. 6, June 1975.

[3] B. Commentz-Walter, "A string matching algorithm fast on the average," Universities Des Saarlandes, June 1979.

[4] Gupta K., Kotagiri R., and Nath B., "Conditional Random Fields for Intrusion Detection," in Proceedings of the 21st International Conference Advanced Information Networking and Applications Workshops, Niagara Falls, pp. 203- 208, 2007.

[5] Gupta K., Nath B., and Kotagiri R., "Layered Approach using Conditional Random Fields for Intrusion Detection," IEEE Transactions on Dependable and Secure Computing, vol. 7, no.1, pp. 35-49, 2010.

[6] ShikhaAgrawal and JitendraAgrawal, "Survey on Anomaly Detection using Data Mining Techniques", 19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems.

[7] Kelton A.P. Costa, "A nature-inspired approach to speed up optimum-path forest clustering and its application to intrusion detection in computer networks", Information Sciences 2014.

[8] D.Dennin,.(1987) "An intrusion-detection model", IEEE Transactions on Software Engineering.

[9] Pfleeger, C. and Pfleeger, S. (2003). Security in computing.Prentice Hall.

[10] M. M. M. Hassan, "Current studies on intrusion detection system, genetic algorithm, and fuzzy logic," International Journal of Distributed and Parallel Systems (IJDPS), 2013.

[11] R. B. Khalid Alsubhi, NizarBouabdallah, "Performance analysis of intrusion detection and prevention systems,"12th IFIP/IEEE International Symposium on Integrated Network Management, 2011.

[12] Sindhu S., Geetha S., and Kannan A., "Decision Tree Based Light Weight Intrusion Detection Using a Wrapper Approach," Expert Systems with Applications, vol. 39, no. 1, pp. 129-141, 2012.

[13] Wilk T. and Michal K., "Soft Computing Methods Applied to Combination of One-class Classifiers," Neuro Computing, vol. 75, no. 1, pp. 185-193, 2012.

[14] Bharat S. Dhak, Shrikant Lade, "An Evolutionary Approach to Intrusion Detection System using Genetic Algorithm", International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 12, December 2012.

[15] B. Kavitha, S.Karthikeyan and B. Chitra,"Efficient Intrusion Detection with Reduced Dimension Using Data Mining classification Methods and Their Performance Comparison", CCIS 70, 2010, pp. 96-101.

[16] KyawThetKhaing (2010),Recursive Feature Elimination (RFE) and k-Nearest Neighbor (KNN) in SVM.

[17] NSL-KDD Data set for Network-based Intrusion Detection Systems.

[18] H. Liu and H. Motoda(1998), Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic.

[19] Pedro Casas, Johan Mazel, Philippe Owezarski, "Unsupervised Network Intrusion Detection Systems: Detecting the Unknown without Knowledge", Computer Communications, Vol. 35, Issue 7, pp. 772-783, 2012

[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] 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.

[22] Sindhu S., Geetha S., and Kannan A., "Decision Tree Based Light Weight Intrusion Detection Using a Wrapper Approach," Expert Systems with Applications, vol. 39, no. 1, pp. 129-141, 2012.

[23] B. Kavitha, S.Karthikeyan and B. Chitra,"Efficient Intrusion Detection with Reduced Dimension Using Data Mining classification Methods and Their Performance Comparison", CCIS 70, 2010, pp. 96-101.

[24] S Aksoy, "Feature Reduction and Selection", Department of Computer Engineering, Bilkent University, 2008, CS 551

[25] Bauer, D. S., &Koblentz, M. E. (1988). NIDX – an expert system for real-time networkintrusion detection

[26] T.Shon, Y. Kim, C.Lee and J.Moon,(2005), A Machine Learning Framework for Network Anomaly Detection using SVM and GA, Proceedings of the 2005 IEEE.

[27] SandyaPeddabachigari, Ajith Abraham, CrinaGrosan, Johanson Thomas (2005). Modeling Intrusion Detection Systems using Hybrid Intelligent Systems.Journal of Network and Computer Applications.

Author's Profile:

G. Krishna Veni, working as an Asst. Professor, in I.T Department, Sir C. R. Reddy College of Engg, Eluru, A.P., India. She has received her B.E (Mech) from S.R.K.R College of Engineering, Bhimavaram, A.P and M.Tech(CSE) from JNTUK,AP. Her research interests include Cloud Computing, Big Data, Data Mining, Networks and Security.

N V S K Vijayalakshmi K, working as an Asst.Prof, in I.T Department, Sir C. R. Reddy College of Engg, Eluru, A.P., India. She has received her B.Tech (CSIT) and M.Tech(SE)from Shri Vishunu Engineering College for Women, Bhimavaram,AP. Her research interests include Cloud Computing, Big Data, Data Mining, Image Processing.

T.Satya Nagamani, working as an Asst. Professor, in I.T Department, Sir C. R. Reddy College of Engg, Eluru, A.P., India. She has received her B.Tech (CSIT) from Lakireddy BaliReddy College of Engineering, Mylavaram,A.P and M.Tech(CSE) from JNTUK, AP. Her research interests include Cloud Computing, Big Data,Data Mining, Networks and Image Processing.