International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-3 Number-3 Issue-12 September-2013

A Review of Intrusion Detection Technique by Soft Computing and Data Mining Approach

Aditya Shrivastava[1], Mukesh Baghel[2], Hitesh Gupta[3]

Abstract

The growth of internet technology spread a large amount of data communication. The communication of data compromised network threats and security issues. The network threats and security issues raised a problem of data integrity and loss of data. For the purpose of data integrity and loss of data before 20 year Anderson developed a model of intrusion detection system. Initially intrusion detection system work on process of satirical frequency of audit system logs. Latter on this system improved by various researchers and apply some other approach such as data mining technique, neural network and expert system. Now in current research trend of intrusion detection system used soft computing approach such as fuzzy logic, genetic algorithm and machine learning. In this paper discuss some method of data mining and soft computing for the purpose of intrusion detection. Here used KDDCUP99 dataset used for performance evaluation for this technique.

Keywords

IDS, Data Mining, soft computing, KDDCUP99

I. Introduction

The current internet technology suffered from a problem of network security and data integrity. For the data integrity and network security various application software are used such as firewall and other scanning antivirus software[1]. The regular monitoring of data and network need power full application and software such as intrusion detection system (IDS). An intrusion detection system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches, which include both intrusions (attacks from outside the organization) and misuse (attacks from within the organization)[5]. IDS uses vulnerability assessment, which is a technology developed to assess the security of a computer system or network. For the increasing size of data and network the ability of intrusion detection and monitoring of file system is compromised [8,9].

Aditya Shrivastava, Department of Computer Science & Engineering PCST, Bhopal, India.
Mukesh Baghel, Department of Computer Science & Engineering PCST, Bhopal, India.
Hitesh Gupta, Department of Computer Science & Engineering PCST, Bhopal, India.

For the improvement of performance of intrusion detection system categories into two section host based intrusion detection system and network based intrusion detection system. The intrusion detection system at present must be able to detect new attacks[21]. Network monitoring systems supervise

the traffic in computer networks and generate alerts and trigger suspicious activity, when suspect activities are detected. The traditional approach to intrusion prevention mostly entails two paradigms, namely, misuse intrusion Detection and anomaly intrusion detection network detection system. Misuse detection systems are most widely used and they detect intruders with known patterns. The signature and pattern used to classify attacks consist of various fields of a network packet, like source address (SA), destination address (DA), source and destination ports or even a number of key words of the payload of a packet[20]. Anomaly detection systems identify deviations from normal behaviour and alert to potential unknown or novel attacks without having any prior knowledge of them. They exhibit higher rate of false prediction but they have the ability of detecting unknown attacks and perform their task of looking for deviations much faster. The elucidation by combining both supervised learning technique and unsupervised learning technique. They used various methods like K Means algorithm for unsupervised learning and Naïve Bayes algorithm for supervised learning. Data mining is about finding insights which are statistically consistent, unidentified previously, and actionable from data .This data must be available, appropriate, satisfactory, and clean. The data mining problems must be well defined and it cannot be solved by query and treatment tools, and guide by a data mining development model. Application and development of specialized machine learning techniques is gaining increasing attention in the intrusion detection community [12,13]. Soft computing is a collection of several methods, which aim to exploit tolerance for indistinctness, uncertainty and incomplete fact to achieve tractability, robustness and low solution, cost. As soft computing techniques can also be used for machine learning, different soft computing techniques have been used for intrusion detection system such as Fuzzy Logic (FL), Artificial Neural Networks (ANN), Genetic Algorithms (GA) and Clustering and outlier detection. Genetic algorithm (GA) field is one of the upcoming fields in computer network security, especially in intrusion detection systems (IDS)[8]. GA operates on a

population of potential solutions applying the principle of survival of the fittest to produce better and better approximations to the solution of the problem that GA is trying to crack. KDD99Cup dataset was found to have quite drawbacks as containing missing and useless features and impossibility of detection of some attacks.KDD99CUP support total 41 features and compute them, some of the features are source IP address, destination IP address,flag,fragment set, services etc. attacks are divided into following categories. Denials-of Service (DoS), Probing, User-to-Root (U2R) and Remote-to-Local (R2L). This paper is divided into five sections. Section-I gives the introduction of the intrusion detection. Section-II gives the related of intrusion detection using soft computing and data mining techniques. Problem formulations in intrusion detection have been reviewed in section-III. In section IV discuss KDDCUP99 data set and empirical evaluation parameter. Finally, in section-V conclusion and future scope.

II. Related Work

In this section discuss related work in current scenario intrusion detection technique using soft computing and data mining approach. In recent research trend soft computing and data mining play a vital role for intrusion detection. The role of data mining such as clustering classification and rule mining apply for detection of known and unknown type of attack, Instead of that soft computing implied in form of attribute and feature selection process in intrusion section system. Some work discuss here in current trend.

[1] In this paper Author applied various neural network classifier methods for analysis the intrusion detection system. Authors used the three types of classifiers used are Feed Forward Neural Network (FFNN), Probabilistic Neural Network (PNN) and Radial Basis Neural Network (RBNN). The feature reduction techniques are used to a given KDD Cup 1999 dataset. The performance of the full featured KDD Cup 1999 dataset is compared with that of the reduced featured KDD Cup 1999 dataset. This study proves that the Probabilistic Neural Networks provides better accuracy over Feed Forward Neural Network and Radial Basis Neural Network.

[2] In this paper author used techniques for protocol type based intrusion detection using neural network .It is experimented that the preprocessing phase plays an important role on the performance of the learning system. It is also observed that applying learning algorithms on divided data (with respect to their protocol types) enables better performance.

[3] In this paper author proposed an Intrusion Detection system (IDS) based Hybrid Evolutionary Neural Network (HENN). In order to construct a precise model for normal behaviors and achieve better detection performance. The genetic algorithm is employed to evolve input features, network structure and connection weights. The experimental results show that the proposed method accomplishes feature selection and structure optimization effectively. Through the comparative analysis, it can be seen that the HENN achieves better detection performance in terms of detection rate and false positive rate.

[4] Author used a Lamster neural network method for intrusion detection, described as Computer systems vulnerabilities such as software bugs are often exploited by malicious users to intrude into information systems. Authors developed an Intrusion Detection System using LAMSTAR neural network to learn patterns of normal and intrusive activities and to classify observed system activities.

[5] In this paper Author proposed an Ensemble Cluster Classification technique using som network for detection of mixed variable data generated by malicious software for attack purpose in host system. In this method SOM network control the iteration of distance of different parameters of assembling.

[6] Author proposed here a using SOM for reduce alarm in IDS and described as an Intrusion detection systems aim to identify attacks with a high detection rate and a low false alarm rate. Classification-based data mining models for intrusion detection are often ineffective in dealing with dynamic changes in intrusion patterns and characteristics. Consequently, unsupervised learning methods have been given a closer look for network intrusion detection. Traditional instance-based learning methods can only be used to detect known intrusions, since these methods classify instances based on what they have learned. They rarely detect new intrusions since these intrusion classes has not been able to detect new intrusions as well as known intrusions. Author proposed a soft Computing technique such as Self organizing map for detecting the intrusion in network intrusion detection. Problems with k-mean clustering are hard cluster to class assignment, class dominance, and null class problems.

[8] In this paper author find unknown or new network attack types with a help of Fuzzy Genetic Algorithm technique. The Fuzzy Genetic Algorithm is rule-based which does not require high computation time. With the obtained detection rule we can detect attacks right after the data arrives. However, it takes about 2 seconds to preprocess the network packets. Therefore, the system requires a total of less than 3 seconds to

issue the alert message after an attack has arrived. From our experiments, the Fuzzy Genetic Algorithm can detect known attack types with high accuracy and low false positive rate which is less than 1%. Moreover, the Fuzzy Genetic Algorithm approach is able to efficiently detect new/unknown attack types with high accuracy. [9]Author proposed here a novel method for IDS using RBFNN and the detailed as a Neural Network model combined with prototype clustering and classification for fast and accurate detection of intrusion in host based system. Previous RBF suffered from grouping of pattern of intrusion, now this problem are reduced using Distance variable ensemble cluster classification and increase the rate of detection of infected data in host system.

[11] In this paper Author proposed a method for intrusion detection based on SARSA and RBF neural network. The proposed method classified attack and normal data of KDDCUP99 is very accurately. The proposed method work in process of making policy of SARSA learning par per rule of policy. The learning process of Q factor and RBF training process makes very efficient classification rate of intrusion data.

III. Problem Formulation In Data Mining And Soft Computing Approach

The soft computing and data mining approach of network intrusion detection system suffered from detection rate and false alarm generation. The process of mining not conformed how many classifier are ensemble for the process of classification of data. The nature of intrusion data is mixed data type but 90% mining technique perform only numerical data for analysis. The conversion of data into one form to another form takes more time and suffered from grouping. The soft computing approach such as neural network and heuristic function compromised with selection of neural network model and appropriate feature selection algorithm for process of classification. Some problem in concern of intrusion detection apply both these technique found in review process [8, 9, 12 20, 22].

1. The pre-processing of KDDCUP99 takes more time.
2. The rate of false alarm generation is high.
3. Some data mining classifier are ambiguous situation for selection of base classifier
4. Entropy based intrusion detection system suffered by high false rate
5. The detection of dynamic feature evaluation as confusion matrix.

IV. Kddcup99 Data Set And Empirical Evaluation Parameter

To check performance of the soft computing and data mining algorithm for intrusion detection and classification, we can evaluate it practically using KDD'99 intrusion detection datasets [1]. In KDD99 dataset these four attack classes (DoS, U2R, R2L, and probe) are divided into 22 different attack classes that tabulated in Table I. The 1999 KDD datasets are divided into two parts: the training dataset and the testing dataset. The testing dataset contains not only known attacks from the training data but also unknown attacks. Since 1999, KDD'99 has been the most wildly used data set for the evaluation of anomaly detection methods. This data set is prepared by[11] and is built based on the data captured in DARPA'98 IDS evaluation program [12]. DARPA'98 is about 4 gigabytes of compressed raw (binary) tcpdump data of 7 weeks of network traffic, which can be processed into about 5 million connection records, each with about 100 bytes. For each TCP/IP connection, 41 various quantitative (continuous data type) and qualitative (discrete data type) features were extracted among the 41 features, 34 features (numeric) and 7 features (symbolic).

Table1: Different types of attacks in kdd99 dataset

4 Main Attack Classes	22 Attack Classes
Denial of Service (DoS)	back, land, neptune, pod, smurt, teardrop
Remote to User (R2L)	ftp_write, guess_passwd, imap, multihop, phf,spy, warezclient, warezmaster
User to Root (U2R)	buffer_overflow, perl, loadmodule, rootkit
Probing(Information Gathering)	ipsweep, nmap, portsweep, satan

To analysis the different results using some standard parameter such as Precision- Precision measures the proportion of predicted positives/negatives which are actually positive/negative. Recall -It is the proportion of actual positives/negatives which are predicted positive/negative. Accuracy-It is the proportion of the total number of prediction that were correct or it is the percentage of correctly classified instances. False-negative rate (FN) is the percentage that attacks are misclassified from total number of attack records. False-positive (FP) is the percentage that normal data records are classified as attacks from total number of normal data records. Below we are showing how to calculate these parameters by the suitable formulas. And also, below we are showing the graph for that particular data set[19].

Precision = $\frac{TP}{TP+FP}$

Recall = $\frac{TP}{TP+FN}$

Accuracy = $\frac{TP+TN}{TP+TN+FN+FP}$

FPR= $\frac{FP}{FP+TN}$, FNR= $\frac{FN}{FN+TP}$

The assessment metrics are computed for testing dataset in the testing phase and the obtained result for all attacks and normal data are given in table 2, which is the overall classification performance of the proposed system on KDD cup 99 dataset. By analyzing the result, the overall performance of the proposed system is improved significantly and it achieves more accuracy for all types of attacks.

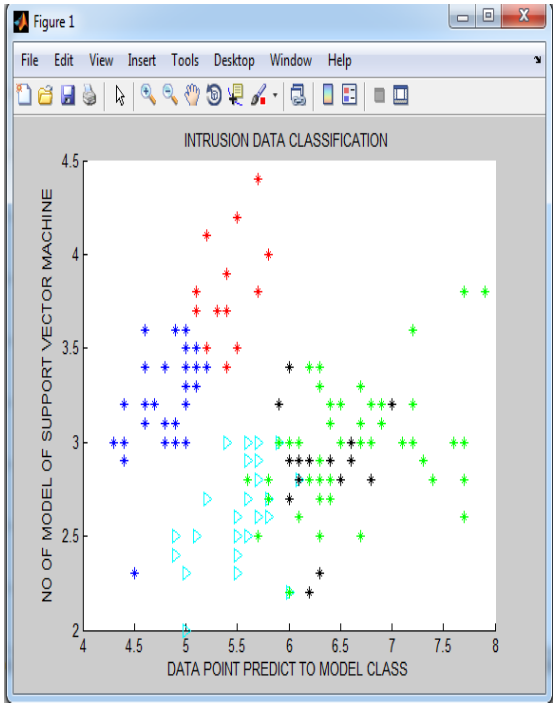

Figure 1: shows that classification process of kddcup99 dataset. In this process used 10,000 data instant in 10000 data instant 7000 instant abnormal data and 3000 instant data of normal data instant.

Table 2: Classification performance of data mining and soft computing algorithm

Kddcup99	algorithm	Accuracy (%)	Precision (%)	Recall (%)
Data-Set	Fuzzy-GA	92.14	87.24	84.43
	SVM	97.14	96.11	94.10
	KNN	89.90	84.32	83.23
	ENSEMBLE	95.23	92.14	91.21
	SOM	91.34	86.14	85.11
	RBF	95.12	93.21	91.13
	HYBRID	92,22	88.21	87.66
	Reinforced learning	97.13	94.52	93.67

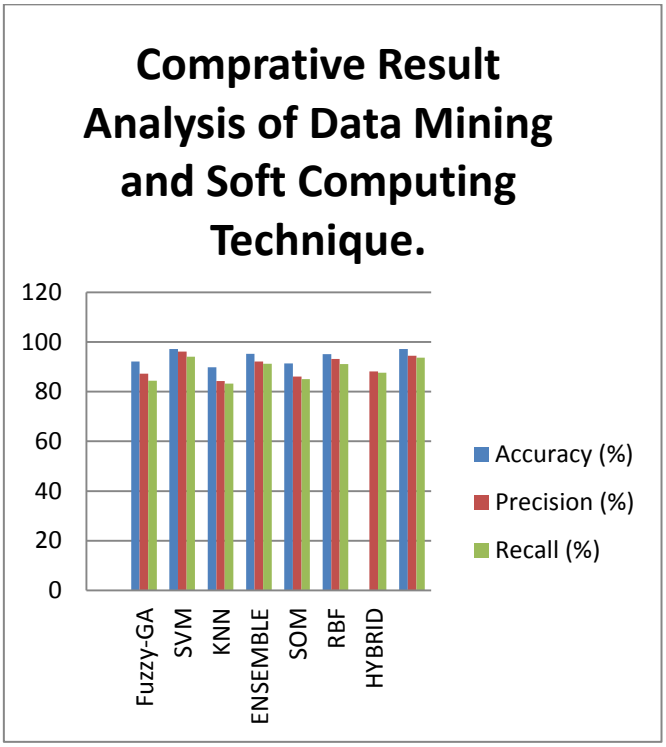

Figure 2: shows that comparative result analysis of soft computing and data mining technique for

V. Conclusion and Future Work

In this paper we review a various method of ensemble classifier and discuss the problem of ensemble classifier for large data. And also discuss the enhancement technique of classifier. Such new ensemble technique is used cluster oriented mechanism for improvement of stream data classification. The selection of optimal number in ensemble classifier is important task. All authors' method suffered from this problem. The selection of ensemble classifier basically based on bagging, boosting and random forest technique. These techniques are not deals in the field of data diversity and suffered stream data classification. For the improvement of data diversity and boundary class training used clustering technique for ensemble classifier. For the survey problem I will solve using ant colony optimization technique for selection of optimal cluster and base boundary value.

References

[1] S.Devaraju and Dr. S.Ramakrishnan "performance analysis of intrusion detection system is using various neural network classifiers" in ieee-international conference on recent trends in information technology, ICRTIT 2011.

[2] Aslıhan Ozkaya and Bekir Karlık "Protocol Type Based Intrusion Detection Using RBF Neural Network" in International Journal of Artificial

Intelligence and Expert Systems (IJAE), Volume (3) : Issue (4) : 2012.
[3] Fan Li "Hybrid Neural Network Intrusion Detection System using Genetic Algorithm" in IEEE 2010.
[4] V.Venkatachalam and S.Selvan "Intrusion Detection using an Improved Competitive Learning Lamstar Neural Network" in IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.2, February 2007.
[5] Deepak Rathore and Anurag Jain "Design Hybrid method for intrusion detection using Ensemble cluster classification and SOM network" in International Journal of Advanced Computer Research Volume-2 Number-3 Issue-5 September-2012.
[6] Singh, Ritu Ranjani, Neetesh Gupta, and Shiv Kumar. "To reduce the false alarm in intrusion detection system using self-organizing map." International Journal of Soft Computing and Engineering (IJSCE) ISSN (2011): 2231-2307.
[7] John Zhong Lei and Ali Ghorbani "Network Intrusion Detection Using an Improved Competitive Learning Neural Network" in Proceedings of the Second Annual Conference on Communication Networks and Services Research IEEE, 2004.
[8] P. Jongsuebsuk,N. Wattanapongsakorn and C. Charnsripinyo "Network Intrusion Detection with Fuzzy Genetic Algorithm for Unknown Attacks" in IEEE 2013.
[9] Deepak Rathore and Anurag Jain "a novel method for intrusion detection based on ecc and radial bias feed forword network" in Int. J. of Engg. Sci. & Mgmt. (IJESM), Vol. 2, Issue 3: July-Sep.: 2012.
[10] Wing w. Y. Ng, rocky k. C. Chang and daniel s. Yeung "dimensionality reduction for denial of service detection problems using rbfnn output sensitivity" in Proceedings of the Second International Conference on Machine Learning and Cybernetics, Wan, 2-5 November 2003.
[11] Chaturvedi, Anshul, and Vineet Richharia. "A Novel Method for Intrusion Detection Based on SARSA and Radial Bias Feed Forward Network (RBFFN)." International Journal of Computers & Technology 7, no. 3 (2013): 646-653.
[12] Mohammad Behdad, Luigi Barone, Mohammed Bennamoun and Tim French "Nature-Inspired Techniques in the Context of Fraud Detection" in ieee transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 42, no. 6, November 2012.
[13] Alberto Fernandez, Maria Jose del Jesus and Francisco Herrera "On the influence of an adaptive inference system in fuzzy rule based classification system for imbalanced data-sets" in Elsevier Ltd. All rights reserved 2009.
[14] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez and E.Vazquez "Anomaly-based network intrusion detection: Techniques, Systems and challenges" in Elsevier Ltd. All rights reserved 2008.
[15] Terrence P. Fries "A Fuzzy-Genetic Approach to Network Intrusion Detection" in GECCO 08, July12–16, 2008, Atlanta, Georgia, USA.
[16] Zorana Bankovic, Dusan Stepanovic,Slobodan Bojanic and Octavio Nieto-Taladriz "Improving network security using genetic algorithm approach" in Published by Elsevier Ltd 2007.
[17] Mrutyunjaya Panda and Manas Ranjan Patra "network intrusion detection using naive bayes" in IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.
[18] Animesh Patcha and Jung-Min Park "An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends" in Computer networks 2007.
[19] Ren Hui Gong, Mohammad Zulkernine and Purang Abolmaesumi "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection" in IEEE 2005.
[20] Jonatan Gomez and Dipankar Dasgupta "Evolving Fuzzy Classifiers for Intrusion Detection" in IEEE 2002.
[21] Francisco Herrera "Genetic fuzzy systems: taxonomy, current research trends and prospects" in Springer-Verlag 2008.
[22] Adel Nadjaran Toosi and Mohsen Kahani "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers" in Elsevier B.V. All rights reserved 2007.

Aditya shrivastava born on 01-12-1990. He received the B.E. degree in information technology with first class from the Patel Institute of Technology, bhopal, India, in 2011.He is presently pursuing final year M.Tech. in computer science and engineering from Patel college of science and technology Bhopal, India.