International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2016 IJSRCSEIT | Volume 1 | Issue 3 | ISSN : 2456-3307

A Data Mining Framework for Intrusion Detection System(IDS) using Bio-Inspired Algorithms

Qamar Rayees Khan[1], Muheet Ahmed Butt[2]
[1]Baba Ghulam Shah Badshah University, Rajouri (J&K), India
[2]University of Kashmir, Srinagar, (J&K), India

ABSTRACT

The information that acts a backbone for any organization is being managed by the sophisticated Information system (IS) is continuously under threat by the vurnabilities by the intrusion. There is always a need to have a better and proper security mechanism in place so that the effect of the intruders may be minimum and at the same time the information may not get compromised, This paper focus on the novel way of dealing the various Intrusions and proposed a model for the Intrusion Detection Systems (IDS). The use of biological inspired algorithms is used in the model for the optimization in terms of efficiency and accuracy. The proposed model comprises of three important phases (Clustering, Mining and classification) to deal with the threats that may occur in the information systems.

Keywords : Information System(IS), intrusion, Intrusion Detection Systems (IDS), biological inspired algorithms, Clustering, Mining and Classification.

I. INTRODUCTION

In the past, money was considered as most important asset in an organization. However, with the passage of time, there was a paradigm shift and the organizations considered information as the most important and valuable asset rather than the financial aspect. The information that is considered as vital in any information system needs a secure mechanism to achieve a higher level of integrity in terms of data and information. For this, a robust information security technique(s) is needed so that the compromise in information integrity is minimized.

The competitive advantage of the business relies on the quality of the information that an organization stores and processes. Any compromise in the nature of the information may lead to the competitive disadvantage as information is considered to be an important ingredient for the policy decision. Various data mining and soft computing techniques have been proposed by the researchers worldwide for detecting intrusions that may compromise the integrity of the information system[1]. However, there always arise a tradeoff between accuracy and consistency of the results that motivates the researchers to continue their research work in this direction.

II. Related Work

Data Mining is one of the discipline of Information Technology that is specifically used to discover the patterns in data[2]. Numerous data mining approaches are used that play significant role in intrusion detection systems. The various contributions of the researchers in this direction are summarized below.

Portnoy et al in their paper proposed an IDS wherein feature vectors were collected and analyzed from the computer networks. The proposed IDS was able to identify new and unknown intrusion attacks [3]

Eskin et al used various data mining techniques like k-Nearest neighbour and allied clustering technique to identify anomalies in network traffic data. [4]

Leung et. al proposed a novel type of algorithm called "fpMAFIA" . It was an unsupervised anomaly detection that works on extra large datasets. The algorithm is density and grid based approach. This dimentional clustering approach shows a rational accuracy rate for detecting anomalies and at the same time maintain a low positive rate [5]

Chittue et. al. used a approach whrein he used a decision tree and for the better results, they used the genetic algorithms. They set the criterion of the approach as the preference criterion by naming the IDS as "Detection rate minus the false positive rate"[6].

Crosbie et. al. proposed a unique approach to identify anomalies detection for which they used sparse trees and genetic algorithms so as to reduce the impact of false positive ratio. [7]

Various other techniques in data mining that does not fit in the classification problems as wee al the clustering problems have shown the significant impact to IDS. J. Zhang et. al. in their paper proposed a state of art technique in data mining for detecting intrusion in networks. The Random Forest technique was used in this paper for malicious activity [8]

Jimmy Shum et. al proposed in their paper an ANN based IDS for detecting malicious activities on computer networks. DARPA date set was used in this paper and the dataset was divided in 70:30 ration for training and testing. The proposed approach showed good results in terms of accuracy but with high false positive ratio.[9]

Usman Ahmed et. al. in their research paper proposed an IDS using Radial-bias-neural networks. The system provide a good intrusion detection rate and comparatively with lesser training time. The approach focus more on time and efficiency apart from minimizing the false rate ratio. [10]

S.Devaraju et. al in their research proposes various classifiers using neural networks so as to evaluate the performance of IDS. They used Matlab with selected features in dataset that shows a significant impact of the intrusion detection rate rather than taking the dataset with all the defined parameters . [11]

Mohd. Junedul et al. in their research contribution in this direction have proposed a clustering approach using k-Means for detection the intrusion . This approach generated massive false positive alarms while detecting the intrusions. [12]

Rohit Arora et. al in their paper perform the comparative analysis of J48 and classification algorithms using Multi-layer Perceptron(MLP). In this paper, they have shown that MLP performed better when compared with J48. [13]

Sneha Kumari et al in their paper used the various data mining techniques for identifying and correspondingly grouping the attackers based on attack level and type. They studied the attack rate as done by the attackers at different intervals [14]

Cannady et. al. used an artificial Neural Network(ANN) IDS that detects the intrusions at the packet level. The syatem dats stored in the computer is collected/ retrieved and the appropriately classified as per the packet properties. The data is then sent to the ANN for the necessary processing [15]

Rayees et al. in this work on intrusion detection using ABC- based IDS revealed the 96% of accuracy and gives 1.39% and 3.41% higher accuracy than ATDIS and MC-SLIPPER respectively[16].

III. Proposed Model and Methodology

A. Data Mining Framework

The existing IDS cannot achieve a good detection rate that in turn affects the integrity of the information present in the information system. In order to overcome the low detection rate of intrusions, an Biological Inspired Algorithm based Intrusion Detection system (BI-IDS) is proposed in figure1.

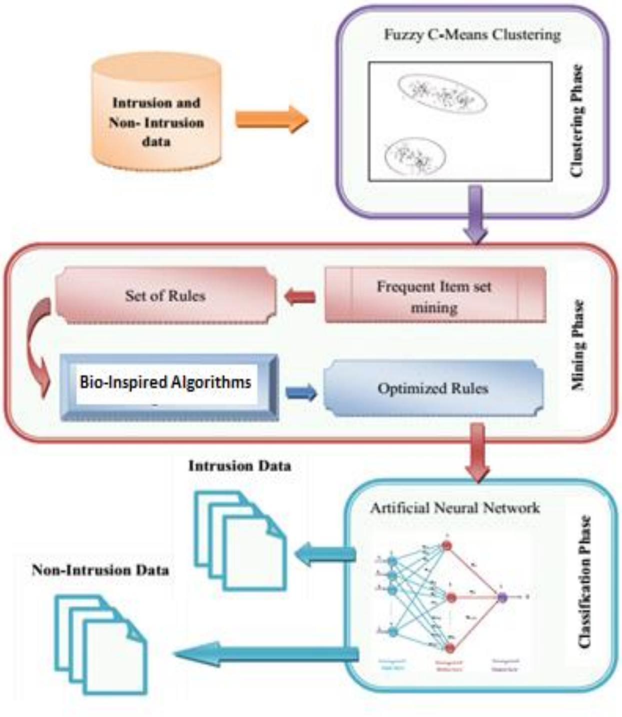

Figure. 1: Proposed Biological Inspired Framework for Intrusion Detection System

The proposed system works in three major phases – Dataset Clustering, Mining and Classification. The Dataset Clustering is the first step in the system that ensures preservation of the quality of the datasets by partitioning the quantity of the datasets as discrete as possible. For the proposed system, a Fuzzy C-Means algorithm is considered as an effective candidate for the clustering. The output of the first step, clustered datasets, are then subjected to mining by use of Frequent Item set mining, that successfully helps to generate some rules. Biological inspired algorithms optimize the rules for efficiency and accuracy. The optimized results are then classified. Finally the input datasets are classified in the Classification phase by Artificial Neural Network Classifier. It differentiates the input dataset into two classes – Intrusion and Non-Intrusion packets. The tests of this Intrusion Detection System is performed over DARPA dataset and is implemented using MATLAB. Some of the data from the same dataset is used for training at the initial stage and then for testing the capabilities and other parameters regarding the intrusion detection.

The Intrusion packets may be well sub categorized into Probe, Denial of Service, User to remote, and remote to local and the Data attacks. To enhance the reliability and improve the precision with consistency, the system is analyzed for its accuracy to distinguish Intrusion data packets and Non-Intrusion data packets.

B. Biological Inspired Algorithm for Optimization

The drawback of using frequent item set mining for obtaining the mined clustered data, is possibility of production of higher number of rules along with the presence of lots of non-frequent item sets as well. To get away from such non-frequent item sets, we use Biological Inspired Algorithms inside the frequent item set mining for rule optimization.

IV. Research Outcome

The outcome of this proposed model is the IDS that is capable of handling the data efficiently by intelligently optimized the efficiency and accuracy of the results. The final results shall classify the data into intrusions and non-intrusion data.

V. Conclusion

In this paper, various techniques are studied using data mining and Artificial Neural networks for classifying and detecting intrusions and non-intrusion data. The paper focus on the Biological inspired IDS that is capable of classifying the data based on the severity of intrusion data. The model is comprised of three important phases and each phase is supposed to do the requisite job as per the applied algorithm. The proposed model shall be able to quantify the intrinsic integrity attributes of accuracy, reliability and consistency of the data and shall perform better with optimal accuracy than the existing IDS techniques.

VI. REFERENCES

[1]. Manish Joshi, “Classification, Clustering and Intrusion Detection System”, International Journal of Engineering Research and Applications (IJERA), pp.961-964, ISSN: 2248-9622 Vol. 2, Issue 2, Mar-Apr 2012.

[2]. Khan Qamar Rayees et al, “Integrity Model based Intrusion Detection System: A Practical Approach”, International Journal of Computer Applications, Vol 115, No. 10, April 2015.

[3]. Portnoy, L., E. Eskin, and S. J. Stolfo (2001), ”Intrusion detection with unlabeled data using clustering”. In Proc. of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001), Philadelphia, pp. 5-8.

[4]. Eskin, E., A. Arnold, M. Preraua, L. Portnoy, and S. J. Stolfo (2002), ”A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data”. In D. Barbar and S. Jajodia (Eds.), Data Mining for Security Applications. Boston: Kluwer Academic Publishers.

[5]. Leung, K. and Leckie, C. (2005), ”Unsupervised Anomaly Detection in Network Intrusion Detection Using Clusters”. In Proc. Twenty-Eighth Australasian Computer Science Conference (ACSC2005), Newcastle, Australia, pp. 333-342, 2005.

[6]. Chittur, A. (2001),”Model generation for an intrusion detection system using genetic algorithms”. High School Honors Thesis, Ossining High School. In cooperation with Columbia Univ, 2001.

[7]. Crosbie, M. and E. H. Spafford (1995), ”Active defense of a computer system using autonomous agents”. Technical Report CSD-TR- 95-008, Purdue Univ., West Lafayette, IN, 15 February 1995.

[8]. J. Zhang and M. Zulkernine (2006), ”Anomaly Based Network Intrusion Detection with Unsupervised Outlier Detection”. Symposium on Network Security and Information Assurance Proc. of the IEEE International Conference on Communications (ICC), 6 pages, Istanbul, Turkey, June 2006.

[9]. Jimmy Shum and Heidar A. Malki (2008), “Network Intrusion Detection System Using Neural Networks”. Fourth International Conference on Natural Computation in IEEE 2008.

[10]. Usman Ahmed, Asif Masood (2009), “Host Based Intrusion Detection Using RBF Neural Networks”. In IEEE 2009 International Conference on Emerging Technologies.

[11]. S.Devaraju, Dr. S.Ramakrishnan (2011), “Performance analysis of Intrusion Detection System using various Neural Network Classifiers”. In IEEE International Conference on Recent Trends in Information Technology, ICRTIT 2011.

[12]. Mohd. Junedul Haque, Khalid.W. Magld, Nisar Hundewale (2012), “An Intelligent Approach for Intrusion Detection Based on Data Mining Techniques”. In IEEE 2012.

[13]. Rohit Arora, Suman (2012), “Comparative Analysis of Classification Algorithms on Different Datasets using WEKA” in International Journal of Computer Applications (0975 –8887), Volume 54– No.13, September 2012.

[14]. Mrs. Sneha Kumari, Dr. Maneesh Shrivastava (2012), “A Study Paper on IDS Attack Classification Using Various Data Mining Techniques”. In International Journal of Advanced Computer Research, Volume-2 Number-3 Issue-5, September-2012.

[15]. Cannady, J., (1998), “Artificial Neural Networks for Misuse Detection,”. Proceedings, National Information Systems Security Conference (NISSC’98), October, Arlington, VA, pp.443-456.

[16]. Khan Q R et al, “Improving Information Integrity using Artificial Bee Colony based

Intrusion Detection System”, International Journal of Computer Applications, Vol 130, No. 06, November 2015.

Cite this Article

Qamar Rayees Khan, Muheet Ahmed Butt, "A Data Mining Framework for Intrusion Detection System(IDS) using Bio-Inspired Algorithms ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 1 Issue 3, pp. 90-93, November-December 2016. Journal URL : https://ijsrcseit.com/CSEIT217234