Instrusion Detection System Based On Support Vector Machine Using BAT Algorithm

Ms Priya Sharma, Mr Anurag Jain

Abstract— Now these day Computers becomes vital part of everyday life and hence use of internet becomes more and more. Due to internet, computers are becomes vulnerable of different kinds of security threats. Therefore it is required that we need to have efficient security method in order to avoid leakage of important data or misuse of data. This security method is called as Intrusion Detection System (IDS). Since from last two decades IDS becomes core area of many researchers and many methods are already presented for efficient intrusion detection and classification. Most of methods are out dated as many new attacks generated by hackers. In our project the main aim is to presented scalable and efficient method for intrusion detection and classifications. Evolutionary algorithm has recently been applied to the anomaly based intrusion detection in computer networks. Evolutionary algorithm is a new technique used to solve various problems in the field of information security. To overcome these deficiencies of the IDS, the network system, a new double detection of IDS based on the integration of Evolutionary algorithm BAT and SVM .The BAT-SVM helps us solve the problem and the correlation theory is proposed model solves the problem of the unknown and the rapid development of damaging attacks.

Index Terms—Intrusion detection systems (IDS) , Artificial immune system (AIS) , Dendritic cell algorithm (DSA) , Support Vector Machine (SVM) , Dempster-belief (DBT), Negative selection algorithm (NSA)

1 INTRODUCTION

Intrusion detection is an effective process in order to monitor the various events happen in a network or individual host [5]. The motive of an IDS is to analyze the traffic of network or host that goes through the IDS by which the detection of intrusions might possible in the system. The design of Intrusion Detection System (IDS) with high efficiency has become much more challenging (16). It is very important to discover abnormal behaviours at early stage, therefore, compared to the traditional signature-based detection, research on anomaly detection has been more popular in academia, as it has the potential power to find unknown attacks by gracious of heuristic learning on the historical training data. Anomaly detection generally contains two steps, constructing a model on training data and using the model for finding. However, training data are usually in a large scale, which can severely delay the detection therefore many detection models may need to scan all of them in certain cases. To address this problem, our work will pay attention largely to the building of data concentration includes of the detection phase, striving to boost the detection efficiency based on a proposed compressed model of training data. As for how to build compressed model, our proposal is create through inspecting into the below common natures of the training data, that is to say the motivation and illumination of our works are generated from the following observations. There are basically two types of intrusion detection system[6]

- **Host-based intrusion detection system.** HIDSs estimate the different information on a single host or multiple hosts of a network which includes operating systems and application files.
- **Network based Intrusion Detection:** NIDS evaluate the information captured from the communication network, analyze the flow of packets moving through the network.

In order to overcome the limitations of these two conventional intrusion detection methods, hybrid intrusion detection methods that combine the misuse detection method and the anomaly detection method have been proposed.

2 RELATED WORK

Varun, C., Arindam, B., Vipin, K [15], this survey attempts to provide a comprehensive and structured overview of the existing research for the problem of detecting anomalies in discrete/symbolic sequences. The aim is to provide a global understanding of the sequence anomaly finding problem and how previous techniques relate to each other. The key grouping of this survey is the categorized of the existing research into three different types, based on the problem formulation that they are trying to solve. These problem formulations are: 1) detecting anomalous orders with respect to a database of normal sequences; 2) detecting an anomalous subsequence between long sequences; and 3) detecting a pattern in a sequence whose frequency of occurrence is anomalous. They display how each of these problem formulations is characteristically different from each other and discuss their related in various application domains. They review techniques from many dissimilar and disconnected application domains that address each of other formulations. Within each problem formulation, our group techniques into typed based on the nature of the dependent algorithm. For each category, they provide a basic anomaly detection technique, and show how the existing techniques are alternative of the basic technique.

Frey, B.J., Dueck, D[16], Clustering data by identifying a subset of representative examples is important for processing sensory signals and detecting patterns in data.

Such "exemplars" can be found by randomly choosing an initial subset of data points and then iteratively refining it, but this works well only if that first choice is close to a good solution. They devised a method called "affinity propagation," which takes as input measures of similarity between pairs of data points. Real-valued messages are interchange between data points up to a high-quality set of exemplars and comparative clump gradually emerges. We used affinity propagation to cluster images of faces, detect genes in microarray data, detecting representative sentences in this manuscript, and detect cities that are efficiently accessed by airline travel. Affinity propagation found clusters with much lower error than other methods, and it did so in less than one-hundredth the amount of time.

Davis J J, Clark A J. Data [18], Data pre-processing is widely recognized as an important stage in anomaly detection. This paper reviews the data pre-processing techniques used by anomaly-based network intrusion detection systems (NIDS), boil down on which aspects of the network traffic are analyzed, and what feature creation and selection methods have been used. Inspiration for the paper comes from the large crack data pre-processing has on the accuracy and capacity of anomaly-based NIDS. The review search that many NIDS limit their view of network traffic to the TCP/IP packet headers. Time-based statistics can be divided from these headers to detect network scans, network worm behaviour, and denial of service attacks. A number of other NIDS perform deeper examination of request packets to find attacks against network services and network applications. Today's approaches analyze full service responses to find attacks targeting clients.

R. Goel, A. Sardana, and R. C. Joshi [13]. In order to achieve high capability of arrangement in intrusion detection, a compressed model is proposed in this paper which binds horizontal compression with vertical compression. One R is utilized as horizontal compression for attribute reduction, and AP is used as vertical consolidation to select small illustrative exemplars from large training data. so it can computationally compress the larger volume of training data with scalability, Map Reduce based parallelization approach is then implemented and calculated for each step of the model compression process above mentioned, on which similar but efficient classification methods can be directly used.

W. Ren, L. Hu, K. Zhao[14] The Internet connects hundreds of millions of computers beyond the world running on many hardware and software platforms providing communication and commercial services. However, this inter connectivity among computers also start malicious users to misuse resources and mount Internet attacks. The continuously develop Internet attacks pose severe challenges to develop an adaptable adaptive security oriented methods. Intrusion detection system (IDS) is one of most valuable component is used to find the Internet attacks. In literature, different techniques from different disciplines have been used to develop efficient IDS. Artificial intelligence (AI) dependant techniques plays prominent role in development of IDS and has many benefits over other techniques. However, there is no comprehensive review of AI based techniques to observe and understand the current status of these techniques to solve the intrusion detection problems. In this paper, many AI based techniques have been reviewed focusing on elaborate of IDS. Related studies have been associate by their source of audit data, processing criteria, technique used, dataset, classifier design, feature reduction technique employed and other experimental environment setup Advantages and Disadvantages of AI based techniques have been discussed.

3 PROBLEM FORMULATION

The enormous growth of computer network increasing the importance of network security.The central challenge with computer security is to develop systems which have the ability to correctly identify an intrusion which represents potentially harmful activity. Therefore, the role of IDS is as special-purpose devices to detect and prevent the anomalies and illegal access of data.In current scenario, users look for the complete security of data at any cost, since security of data become prime requirement for everyone.The new challenge requires several changes in existing IDS system in order to improve the correlation of alarm; the detection and prediction of false positive and false negative rate must be low. Recently, using biological models such as neural networks and genetic algorithms in modelling and solving computational problems has been spectacularly successful. Lots of traditional IDS techniques are only able to detect and prevent known intrusions and mostly are static. They are not able to recognize unknown intrusions. The biological models has some features such as self-organized, automated, distributed etc., which are now IDS starve for. So AIStheory for detecting intrusion becomes a new immerging approach in security research.

4 PROPOSED METHODOLOGY

With the increase of malicious network activities, considerable attentions have been paid to intrusion detection system. The network intrusion detection system is designed to classify anomalous behaviours by examining the dynamic characteristics of network connection records; its role is becoming more important as a vital part of the network security architecture.
The central challenge with computer security is developing systems which have the ability to differentiate between the normal and an intrusion which represents potentially harmful activity. A promising solution is emerging in the form of biologically inspired computing, and in particular artificial immune systems (AIS).

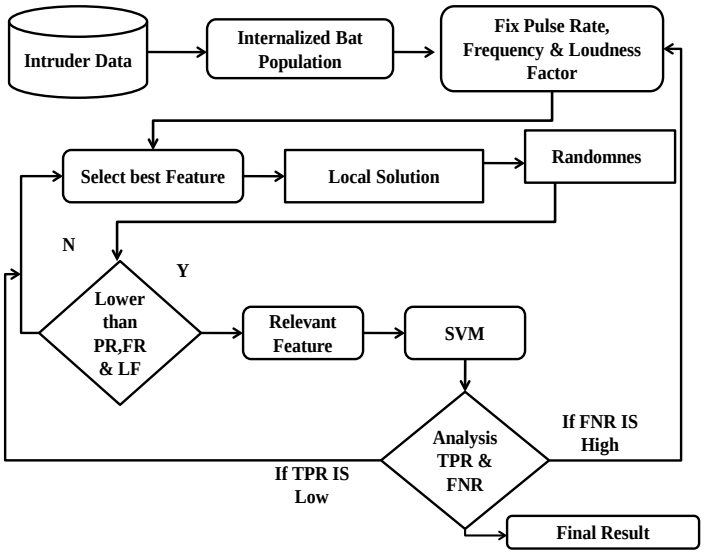

Figure 1 Proposed Frameworks

Proposed introduced a novel intrusion detection system to identify all intrusion correctly and minimized the false alarm generation. Proposed methodology encapsulate BAT along with the concept of SVM is used to designed and increase the performance of the system. SVM scale the uncertainty in feature selected by BAT and on the basis of randomness refine the feature. If feature set have higher entropy it's again send for decision otherwise consider.

Proposed framework initially use intruder data (KDD 99 data set) for training then generate randomized feature for IDS system. In proposed framework initially intruder data set initialized with random partition. Then BAT approach tends to generate random feature for intruder. This BAT is recertifying through randomness. If randomness of relevant feature is low then it's acceptable otherwise feature is not to be consider. Acceptable Feature is denoted as relevant feature and apply for classify intruder. If intruder detection have high false negative rate then whole process is initialed by random partition and if intruder detection have low true positive rate then new feature is generated on same partition.

This paper use proposed technique for intrusion detection that is based on one of the evolutionary algorithm known as BAT and apply SVM for classification. With the help of BAT we calculate the degree of uncertainty and again on the basis of randomness with the classify the end user, end user having higher entropy, is regarded as the "intruder", and generate the alarm. Proposed System can not only reduce the false positive and false negative rate but also improved the correlation technique and decrease the intrusion rate in the system.

Experimental Setup & Result analysis

Proposed Work has implemented in JAVA Netbeans framework. The result of classification of proposed IDS is shown in Table 1. In the proposed IDS the accuracy for the classification of data for generating function 0.9 reaches up to 94.309% with minimum FPR and FNR whereas classification rate for Existing Method (IG-ABC SVM) .In ABC method the accuracy for the classification of data for generating function 0 .9 is 89.799%. Proposed methodology is very effective for the classification of data with maximum accuracy and minimum FPR and FNR.

Table 1 Comparison between Accuracy Rates

Generating value	IG-ABC-SVM	IG-BAT-SVM
0.2	91.40	95.90
0.3	89.79	94.30
0.5	91.55	96.06
0.7	91.53	96.04
0.8	91.55	96.06
0.9	89.79	94.30
0.1	89.29	94.30

Figure 2 shows comparison of the simulation result .It gives the comparison of the degree of Accuracy rate of IDS system by using traditional classification method namely SVM with our proposed method IG-BAT-SVM. IG-BAT-SVM increases the accuracy rate by encapsulating BAT along with SVM method.

As shows in figure 2 SVM classification algorithm alone having accuracy rate for attack detection never reaches even 92.00% whereas IG-BAT-SVM model having accuracy rate up to 96.00%. The X-axes represents the accuracy rate and the Y-axes indicate detection generating value.

Figure 3 shows comparison of the true positive rate of IDS system by using traditional classification method namely BAT with our proposed method. In Proposed modal because of higher degree of filtering minute suspicious data take as abnormal data that's leads to minimizing the true positive rate by encapsulating BAT along with SVM method .As shows in figure 3 SVM classification algorithm alone having higher level of true positive rate because of lower level of filtering whereas in proposed model due to multilevel filtering having lower level of true positive rate. The X-axes represents the rate and the Y-axes indicate detection generating value.

Figure 4 also shows comparison of the true negative rate of IDS system by using traditional classification method namely SVM with our proposed method. Here same as true positive rate because of multilevel filtering or verification proposed modal having lower true negative rate .As shows

in figure 4 SVM classification algorithm alone having higher level of true nigative rate because of lower level of filtering whereas in proposed model due to multilevel filtering having lower level of true negative rate. False positive means if any data is abnormal and our system take it as normal ,higher FPR leads lower level of accuracy. Figure 4 also shows comparison of the false positive rate of IDS system.

Table2 Comparison between Performance Evaluations

Generating Value	Method	TPR	TNR	FPR	FNR
0.2	IG-ABC-SVM	5.91	3.25	3.05	2.85
0.2	IG-BAT-SVM	1.65	2.5	2.6	2.69
0.3	IG-ABC-SVM	4.27	1.61	1.41	1.21
0.3	IG-BAT-SVM	0.008	0.868	0.966	1.055
0.5	IG-ABC-SVM	6.03	3.37	3.17	2.97
0.5	IG-BAT-SVM	1.76	2.62	2.72	2.81
0.6	IG-ABC-SVM	4.27	1.61	1.41	1.21
0.6	IG-BAT-SVM	0.008	0.868	0.966	1.055

By using traditional classification method namely random forest with our proposed method. As per requirement proposed modal minimizing the false positive rate by encapsulating BAT along with SVM method that's trend to lead higher accuracy rate .

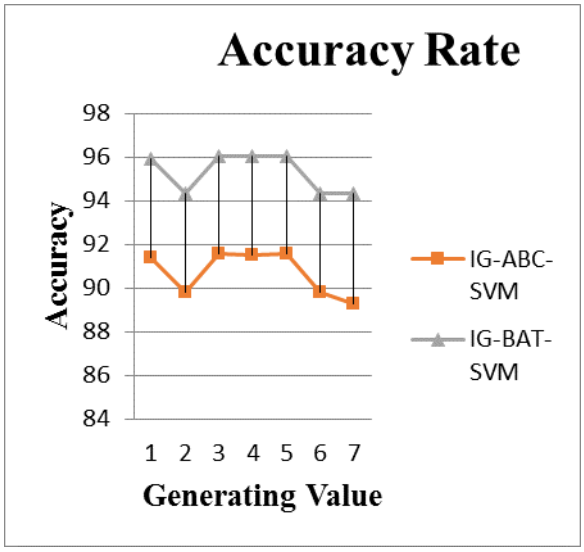

Figure 4: Comparison graph

False negative means if any data is normal and our system take it as abnormal. Figure 5 also shows comparison of the false negative rate of IDS system by using traditional classification method namely SVM with our proposed method. Here same as true positive rate and true negative rate because of multilevel filtering or verification Proposed modal having lower false negative rate .As shows in figure 5SVM classification algorithm alone having higher level of false negative rate because of lower level of filtering whereas in proposed model due to multilevel filtering having lower level of false negative rate.

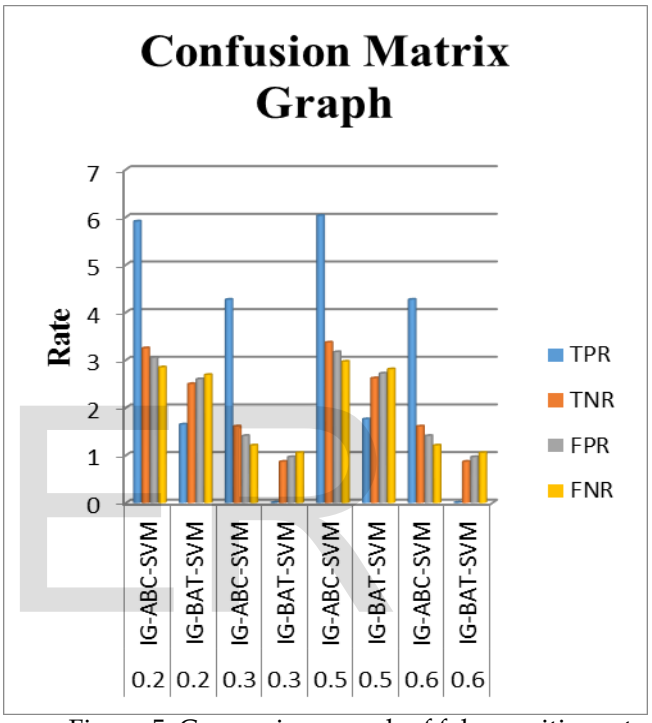

Figure 5: Comparison graph of false positive rate

This chapter contain simulation detail and experiment results which shows the proposed method has improve the correlation factor, minimizing the false +ve and false –ve alarm generation and, to increase the rate of detection of intrusion. So it is a better solution of Intrusion detection the feature reduction process of KDD dataset takes large amount of time.

5 Conclusion

As rapid increase in unauthorized activities and abuse of computer system by both system insider and external intruder trends to increase the degree of network security. In order to increase network security various technique has been proposed but having a deficiency over IDS system in some of the situation i.e. if correlation alarm is not precise, reduction and prevention of false positive and false negative is high , at last having insufficient measurement of pattern recognition. In order to overcome all these deficiency from IDS, system over network ,we propose a novel dual detection of IDS based on evolutionary algorithm that inte-

grating the BAT and SVM .The BAT helps us to solve the problem of correlation and SVM theory resolves the problem of unknown and rapidly evolving harmful attacks. The simulation results shows that the proposed method has improved the correlation factor, minimizing false +ve and false –ve alarm generation and to increase the efficiency and accuracy of the IDS system.

However, the feature reduction process of KDD dataset takes large amount of time. Therefore in future work for modify feature reduction optimization for the better selection of feature in KDD dataset can be attempted.

REFERENCES

[1] S. X. Wu and W. Banzhaf, "The Use of Computational Intelligence in Intrusion Detection Systems: A Review," Applied Soft Computing, vol. 10, 2010, pp. 1–35.
[2] PeymanKabiri and Ali A. Ghorbani, "Research on Intrusion Detection and Response: A Survey", International Journal of Network Security, vol. 1, 2005, pp. 84-102.
[3] C. F. Tsai, Y. F. Hsu, C. Y. Lin, and W. Y. Lin, "Intrusion Detection by Machine Learning: A Review," Expert Systems with Applications, vol. 36, December 2009, pp. 11994–12000.
[4] K. Shafi and H. A. Abbass, "An Adaptive Genetic-based Signature Learning System for Intrusion Detection," Expert Systems with Applications, vol. 36, 2009, pp. 12036–12043.
[5] A. Ahmed, A. Lisitsa, and C. Dixon, "A Misuse-Based Network Intrusion Detection System Using Temporal Logic and Stream Processing," Proc. International Conference on Network and System Security (NSS 11), 2011, pp. 1 – 8.
[6] S. Petrovic and K. Franke, "A New Two-Stage Search Procedure for Misuse Detection," Proc. International Conference on Future Generation Communication and Networking (FGCN 07), 2007, pp. 418 – 422.
[7] P. G. Teodoro, J. D. Verdejo, G. M. Fernández, and E. Vazquez, "Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges," Computers & Security, vol. 28, 2009, pp. 18–28.
[8] Jonathan J. Davis and Andrew J. Clark, "Data Preprocessing for Anomaly based Network Intrusion Detection: A Review", Computers & Security, vol. 30, 2011, pp. 353–375.
[9] F. Palmieri and U. Fiore, "Network Anomaly Detection through Nonlinear Analysis," Computers & Security, vol. 29, October 2010, pp. 737–755.
[10] M. A. Aydın, A. H. Zaim, and K. G. Ceylan, "A Hybrid Intrusion Detection System Design for Computer Network Security," Computers & Electrical Engineering, vol. 35, May 2009. pp. 517–526.
[11] C. Xiang, P. C. Yong, and L. S. Meng, "Design of Multiple-Level Hybrid Classifier for Intrusion Detection System using Bayesian Clustering and Decision Trees," Pattern Recognition Letters, vol. 29, May 2008, pp. 918–924.
[12] P. Gogoi1, D. K. Bhattacharyya, B. Borah, and J. K. Kalita, "MLH-IDS: A Multi-Level Hybrid Intrusion Detection Method," The Computer Journal, vol. 57, 2013, pp. 602-623.
[13] R. Goel, A. Sardana, and R. C. Joshi, "Parallel Misuse and Anomaly Detection Model," International Journal of Network Security, vol. 14, July 2012, pp. 211-222.
[14] W. Ren, L. Hu, K. Zhao, and J. Chu, "A Multiple-Level Hybrid Intrusion Detection System based on Hierarchical Clustering and Decision Trees," Journal of Computational Information Systems, vol. 9, 2013, pp. 5421–5428.
[15] Varun, C., Arindam, B., Vipin, K.: Anomaly Detection, A Survey. ACM Computing Surveys, 2009,41(3):1-58.
[16] Frey, B.J., Dueck, D. Clustering by Passing Messages between Data Points. Science, 315(5814), 972-976 (2007)
[17] Davis J J, Clark A J. Data pre-processing for anomaly based network intrusion detection. A review [J]. Computers and Security, 2011, 30(6): 353-375.

IJSER