Research Article Impact Factor: 0.621 ISSN: 2319-507X
Naved Raza Q Ali, IJPRET, 2014; Volume 2 (9): 424-433 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

A SURVEY ON CLOUD BASED INTRUSION DETECTION SYSTEM USING ARTIFICIAL NEURAL NETWORKS AND FUZZY CLUSTERING

NAVED RAZA Q. ALI[1], PROF. KISHOR B. SADAFALE[2]

1. Pursuing the master's degree, I.T. Department, Sinhgad College of Engineering, University of Pune.
2. Assistant Professor, I.T. Department, Sinhgad College of Engineering, University of Pune.

Accepted Date: 27/02/2014 ; Published Date: 01/05/2014

Abstract: The world is becoming more interconnected with the advent of the Internet and new networking technology. There is a large amount of personal, commercial, military, banking, government and other important data on networking infrastructures worldwide. To protect these data from various attacks, hackers, intrusions, malware, DDoS attacks or disgruntled employees one of the popular components in network security is used, known as Intrusion Detection Systems (IDS). IDS are commonly, software that automates the intrusion detection process and detects possible intrusions. Artificial Neural Networks (ANNs) provide the potential to identify and classify network activity based on limited, incomplete, and nonlinear data sources. According to many researches, Artificial Neural Networks can improve the performance of intrusion detection systems (IDS) when compared with traditional methods. However for ANN-based IDS, detection precision, especially for low-frequent attacks, and detection stability are still needed to be enhanced. To improve the learning capabilities and reduce the computational intensity of competitive learning comparing the performance of the algorithms is performed respectively, different dimension reduction techniques have been proposed. This paper provides a review on current trends in intrusion detection together with a study on technologies implemented by some researchers in this research area. This paper presents the IDS system with ANN, Fuzzy clustering and other techniques for intrusion detection.

Keywords: Instruction Detection System, Artificial Neural Networks, SVM, Fuzzy C-Means clustering, KDDCup 1999, Network Security

Corresponding Author: MR. NAVED RAZA Q. ALI

Access Online On:

www.ijpret.com

How to Cite This Article:

Naved Raza Q Ali, IJPRET, 2014; Volume 2 (9): 424-433

PAPER-QR CODE

INTRODUCTION

With the rapid development and expansion of network technologies and applications, network attacks are greatly increasing both in number and severity. As a key technique in network security domain, Intrusion Detection Systems (IDS) plays vital a role of detecting various kinds of attacks and provides the security over the networks. Main purpose of IDS is to find out intrusions among normal audit data and this can be considered as a classification problem. Intrusion detection systems (IDS) are an effective security technology, which can detect, prevent and possibly react to the attack. It performs monitoring of target sources of activities, such as audit and network traffic data in computer or network systems, requiring security measures, and employs various techniques for providing security services. With the tremendous growth of network based services and sensitive information on networks, network security is becoming more and more important than ever before [12].

IDSs can be divided into two classes:

(1) network-based and (2) host-based. Network intrusion detection systems (NIDSs) analyses network packets captured from a network segment, while host-based intrusion detection systems (HIDSs) such as IDES (Intrusion Detection Expert System) [18] examine audit trails or system calls generated by individual hosts.

1.1 Types of Intrusion Detection System (IDS)

IDSs can also be divided according to the detection approaches they use. There are two detection methods: misuse detection and anomaly detection. The main deference between these two methods is that the misuse detection identifies intrusions based on features of known attacks while anomaly detection analyzes the properties of normal behavior. The two detection approaches explain by following subsections.

1.1.1 Misuse Detection System

Misuse detection identifies intrusions in terms of the characteristics of known attacks. Any action that conforms to the pattern of a known attack or vulnerability is considered as intrusive. The main issues in misuse detection system are how to write a signature that encompasses all possible variations of the pertinent attack. And how to write signatures that do not also match non-intrusive activity. Block diagram of misuse based detection system is as following.

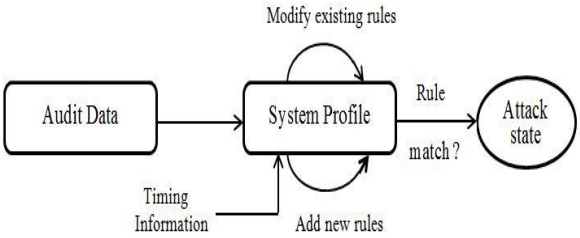

Fig 1.1 Misuse Detection Systems

Misuse detection identifies intrusions by matching monitored events to patterns or signatures of attacks. The attack signatures are the characteristics associated with successful known attacks. The major advantage of misuse detection is that the method possesses high accuracy in detecting known attacks. However, its detection ability is limited by the signature database. Unless new attacks are transformed into signatures and added to the database, misuse-based IDS cannot detect any attack of this type. Deferent techniques such as expert systems, signature analysis, and state transition analysis are utilized in misuse detection.

1.1.2 Anomaly Detection System

Anomaly detection is based on the normal behavior of a subject (e.g. a user or a system). Any action that significantly deviates from the normal behavior is considered as intrusive. That means if we could establish a normal activity profile for the system, then we can flag all system states varying from established profile.

The block diagram of anomaly detection system is as following,

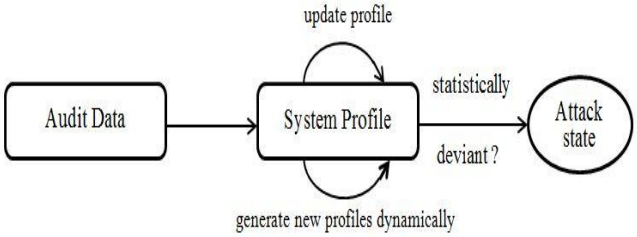

Fig 1.2 Anomaly Detection Systems

There is an important difference between misuse based and anomaly based technique that the misuse based detection system try to recognize the known bad behavior and anomaly based try to detect the compliment of bad behavior. In this case we have two possibilities:

(1) False positive: Anomalous activities that are not intrusive but are flagged as intrusive.

(2) False Negative: Anomalous activities that are intrusive but are flagged as non-intrusive.

Anomaly detection assumes that intrusions are anomalies that necessarily differ from normal behavior. Basically, anomaly detection establishes a profile for normal operation and marks the activities that deviate significantly from the profile as attacks. The main advantage of anomaly detection is that it can detect unknown attacks However; this advantage is paid for in terms of a high false positive rate because, in practice, anomalies are not necessarily intrusive. Moreover, anomaly detection cannot detect the attacks that do not obviously deviate from normal activities. As the number of new attacks increases rapidly, it is hard for a misuse detection approach to maintain a high detection rate. In addition, modeling attacks is a highly qualified and time- consuming job that leads to a heavy workload of maintaining the signature database [18]. On the other hand, anomaly detection methods that discover the intrusions through heuristic learning are relatively easy to maintain.

Paper Organization: Section I has introduced the Intrusion Detection System, basic ideas in intrusion detection and the motivations for this study. In Section II, we discussed the related works on IDS. Section III presents the artificial neural networks theory and its types. Section IV presents the SVM system or Support Vector Machine for IDS. Section V presents Fuzzy Logic (IDS-based), which also applies unsupervised learning methods. Section VI presents the Fuzzy C-Means Clustering algorithm. The Section VII is an overview of the four major categories of networking attacks (i.e., DoS, R2L, U2R, and Probing). Section VIII presents KDDCup 1999; this database contains a standard set of data to be audited, which include a wide variety of intrusions simulated in a military network environment. Finally, section IX provides the concluding remarks of the paper.

II. RELATED WORKS

Anderson *et al*. (1980) has introduced Intrusion Detection System and his work has been improved by (Dorothy, 1987). On the basis of their experiment, user behaviors are translated using some computer audit mechanism and other statistical detection methods to detect masqueraders who illegally access the system. In recent years, data mining algorithms have been applied as intrusion detection methods for finding new intrusion patterns (Dorothy, 1987; Nicholas *et al*., 1996; Kusum *et al*., 2010; Al; Mehdi, 2012). Fuzzy C-Means (FCM) based algorithms are the most popular fuzzy clustering algorithms in practice (Jakir *et al*., 2011; Suguna and Selvi, 2012). Classification is a type of supervised learning that is used to classify data into particular category. Under classification methods, there exist a variety of classifiers which have been widely cited, used and reviewed by different researchers such as by (Xindong *et al*., 2008; Chih-Fong *et al*., 2009). Other classifiers have attracted researcher's interest in recent years, such as Support Vector Machine (Vladimir, 1995) and Neural Networks (Sang-Jun,

2005). Neural computing refers to a pattern recognition methodology for machine learning (Anderson, 1995). ANN is the most popular used approaches in IDS (Mehdi, 2011).

Dahlia and Zurina (2013) used the classic feed-forward neural network with back-propagation algorithm or known as Multilayer Perceptron (MLP) to predict intrusion. The KDD dataset was used for anomaly dataset and the result of accuracy of classification was 92.2% using MLP Neural Network and 93.2% using RBF Neural Network.Recently, a lot of learning techniques have been explored in clustering and classification for the task of anomaly detection, for example, as studied by (Wang *et al*., 2010; Muna *et al*., 2012). Hybrid concept is gaining popularity as the approach promises better flexibility in detecting malicious traffic.

III.ARTIFICIAL NEURAL NETWORKS

An Artificial Neural Networks (ANNs) is an information processing system that is inspired by the way biological nervous systems, such as the brain, process information. It is composed of a large number of highly interconnected processing elements (neurons) working with each other to solve specific problems. Each processing element (neuron) is basically a summing element followed by an activation function. The output of each neuron (after applying the weight parameter associated with the connection) is fed as the input to all of the neurons in the next layer [13].

Advantages of Artificial Neural Networks:

1. It has self-learning capability.

2. Performs tasks that a linear program cannot.

3. When an element of the neural network fails, it can continue without any problem due to their parallel nature.

4. A neural network learns and does not need to be reprogrammed.

The goal for using ANNs for intrusion detection is to be able to generalize from incomplete data and to be able to classify data as being normal or intrusive. Detection precision and detection stability are two key indicators to evaluate intrusion detection systems [22] that refer to the accuracy for each class of attack and stability the detection in each class respectively. Recently, there has been exhaustive effort in improving the existing anomaly detection techniques due to significantly high false alarm, as well as moderate accuracy and detection rate. In addition, there is lacking in performance of a single classifier, which has resulted in high tendency for wrong classification while detecting unknown attacks [5].

3.1 Types of ANN based Intrusion Detection Systems

In specific, there are three types of ANN based IDS [5].

3.1.1 Supervised ANN-based IDS:

It forms input-output pair examples to build an external relationship between the input and output. But since in practice the number of training set is very large and the distribution of training set is imbalanced, the MLFF neural networks is easy to reach the local minimum and thus stability is lower.

3.1.2 Un Supervised ANN-based IDS:

It classify input data and separate normal behaviors from abnormal or intrusive ones. The main advantage of unsupervised ANN in IDS is that it can improve the analysis of new data without retraining. Just like using supervised learning ANN, the performance of unsupervised ANN is also lower. Especially for low frequent attacks, unsupervised ANN also gets lower detection precision.

3.1.3 Hybrid ANN-based IDS:

It can be formed by combining supervised ANN and unsupervised ANN, or combine the ANN with other data mining techniques to detect intrusion. In paper [9] author introduced a combination of SOM and Radial Basis Function (RBF) networks. The system offers generally better results than IDS based on RBF networks alone. In paper [3] author proposed hybrid flexible neural-tree-based IDS based on flexible neural tree, evolutionary algorithm and Particle Swarm Optimization (PSO). Empirical results indicated that the proposed method is efficient. Different ways to construct hybrid ANN will highly influence the performance of intrusion detection. Different hybrid ANN models should be properly constructed in order to serve different aims.

IV. SUPPORT VECTOR MACHINE SVM-IDS

Support Vector Machines or SVMs, are learning machines that plot the training vectors in high dimensional feature space, labeling each vector by its class. SVMs look at the classification problem as a quadratic optimization problem.

They combine generalization control with a method to prevent the “curse of dimensionality” by placing an upper bound on the margin between the different classes, making it a practical tool for large and dynamic data sets. The categorization of data by SVMs is done by determining a set of support vectors, which are members of the set of training inputs that outline a hyper plane in feature space.

There are two main reasons for experimentation with SVMs for intrusion detection. The first is speed because real time performance is of key importance to intrusion detection systems, and any classifier that can potentially outrun neural networks is worth considering. The second reason is scalability: SVMs are relatively insensitive to the number of data points and the classification complexity does not depend on the dimensionality of the feature space [8].

V. FUZZY-IDS

With the fuzzy input sets defined, the next step is to write the rules to identify each type of attack. A collection of fuzzy rules with the same input and output variables is called a fuzzy system. We believe the security administrators can use their expert knowledge to help create a set of rules for each attack. The rules are created using the fuzzy system editor contained in the Matlab Fuzzy Toolbox. This tool contains a graphical user interface that allows the rule designer to create the member functions for each input or output variable, create the inference relationships between the various member functions, and to examine the control surface for the resulting fuzzy system. It is not expected, however, that the rule designer utterly relies on intuition to create the rules. Visual data mining can assist the rule designer in knowing which data features are most appropriate and relevant in detecting different kinds of attacks [8].

VI. FUZZY C-MEANS CLUSTERING

Fuzzy C-means Clustering (FCM) is a clustering technique which is separated from hard k-means that employs hard partitioning. FCM is an iterative algorithm. The aim of FCM is to find cluster centers (centroid) that minimize a dissimilarity function. This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point [7]. Fuzzy Clustering also called soft clustering. In fuzzy clustering we make a fuzzy partition of the data set.

VII. NETWORKING ATTACKS

This section is an overview of the four major categories of networking attacks. Every attack on a network can comfortably be placed into one of these groupings [15].

1. Denial of Service (DoS):

A DoS attack is a type of attack in which the hacker makes a computing or memory resources too busy or too full to serve legitimate networking requests and hence denying users access to a machine e.g. apache, smurf, neptune, ping of death, back, mail bomb, UDP storm etc. are all DoS attacks.

2. Remote to User Attacks (R2L):

A remote to user attack is an attack in which a user sends packets to a machine over the internet, which s/he does not have access to in order to expose the machines vulnerabilities and exploit privileges which a local user would have on the computer e.g. xlock, guest, xnsnoop, phf, send mail dictionary etc.

3. User to Root Attacks (U2R):

These attacks are exploitations in which the hacker starts off on the system with a normal user account and attempts to abuse vulnerabilities in the system in order to gain super user privileges e.g. perl, xterm.

4. Probing:

Probing is an attack in which the hacker scans a machine or a networking device in order to determine weaknesses or vulnerabilities that may later be exploited so as to compromise the system. This technique is commonly used in data mining e.g. saint, portsweep, mscan, nmap etc.

VIII. KDD CUP 1999 DATA

This was the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99, the Fifth International Conference on Knowledge Discovery and Data Mining. The main objective of the competition was to build a network intrusion detector, a predictive model capable of making a distinction between bad connections, called intrusions or attacks, and good normal connections. This database contains a standard set of data to be audited, which include a wide variety of intrusions simulated in a military network environment. Most of the intrusion detection systems since 1999 have been tested and trained on this dataset [8].

IX. CONCLUSION

Intrusion detection systems (IDSs) play an important role in computer security. IDS users relying on the IDS to protect their computers and networks demand that IDS provides reliable and continuous detection service. ANN based IDS provides highly accuracy, reliability and fast detection. Fuzzy C-Means algorithm with IDS provides high stability to detect various types of network intrusions. To implement and measure the performance of IDS system the standard KDD99 benchmark dataset is used and obtained reasonable detection rate.

REFERENCES

1. Anderson, J.P, “Computer security threat monitoring and surveillance”, 1980.

2. Anderson, J.P, "An introduction to neural networks" 3rd Edn., MIT Press, Cambridge, ISBN-10: 0262011441, pp: 672,1995.

3. Chen, Y., A. Abraham and B. Yang, "Hybrid flexible neural-tree-based intrusion detection systems", Int. J. Intell. Syst., 22: 337-352. DOI: 10.1002/int.20203, 2007.

4. Chih-Fong, T., H. Yu-Feng, L. Chia-Ying and L. Wei Yang, "Intrusion detection by machine learning: A review", Expert Syst. Appli., 36: 11994- 12000. DOI: 10.1016/j.eswa.2009.05.029, 2009.

5. Dahlia Asyiqin Ahmad Zainaddin and Zurina Mohd Hanapi, "HYBRID OF FUZZY CLUSTERING NEURAL NETWORK OVER NSL DATASET FOR INTRUSION DETECTION SYSTEM" Journal of Computer Science, 9 (3): 391-403, ISSN 1549-3636, Department of Industrial Electronic, Faculty of Computer and Network Technology, German Malaysian Institute, Malaysia 2013.

6. Dorothy, E.D, "An intrusion-detection model" IEEE Trans. Software Eng., 13: 222-232. DOI:10.1109/TSE.1987.232894, 1987.

7. Esh Narayan, Pankaj Singh, "Intrusion Detection System Using Fuzzy C_ Means Clustering with Unsupervised Learning via EM Algorithms" VSRD-IJCSIT, Vol. 2 (6), 502-510, Research Scholar, 1,3Department of Computer Science & Engineering, 2Department of Electronics & Communication Engineering, Lovely Professional University, Jalandhar, Punjab, INDIA, 2012.

8. Hamdan.O.Alanazi, Rafidah Md Noor, B.B Zaidan, A.A Zaidan. "Intrusion Detection System: Overview" JOURNAL OF COMPUTING, VOLUME 2, ISSUE 2, ISSN 2151-9617, FEBRUARY 2010.

9. Horeis, T, "Intrusion detection with neural network-Combination of selforganizing maps and redial basis function networks for human expert integration", 2003.

10. Jakir, H., A. Rahman, S. Sayeed, K. Samsuddin and F. Rokhani, "A modified hybrid fuzzy clustering algorithm for data partitions", Australian J. Basic Applied Sci., 5: 674-681, 2011.

11. Kusum, K., S.S. Bharti and S. Jain,"Intrusion detection system using clustering", Proceedings of the International Conference, (IC' 10), 2010.

12. Meera Gandhi and S.K.Srivatsa, "Detecting and preventing attacks using network intrusion detection systems" International Journal of Computer Science and Security, Volume (2): Issue (1), Department of Computer Science and Engg., Research Scholar, Sathyabama University.

13. Mehdi, M. and M. Zulkernine, "A Neural Network Based System for Intrusion Detection and Classification of Attacks", University of British Columbia, 2011.

14. Mehdi, B. and B. Mohammad, "An overview to software architecture in intrusion detection system" Int. J. Soft Comput. Software Eng. DOI: 10.7321/jscse.v1.n1.1, 2012.

15. Mohammad Sazzadul Hoque, Md. Abdul Mukit "An implementation of intrusion detection system using genetic algorithm" International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2, DOI 10.5121/ijnsa.2012.4208 109, March 2012.

16. Muna, M., T. Jawhar and M. Mehrotra, "A hybrid FCM clustering-neural network for intrusion detection" Department of Computer Science, Jamia Millia Islamia New Delhi, 2012.

17. Nicholas, Zhang, M. Chung, B. Mukherjee and R.A," A Methodology for testing Intrusion Detection System", IEEE Trans. Software Eng., 22: 719-729. DOI: 10.1109/32.544350, 1996.

18. Poonam Dabas and Rashmi Chaudhary , "Survey of Network Intrusion Detection Using K-Mean Algorithm" International Journal of Advanced Research in Computer Science and Software Engineering 3(3), pp. 507-511, Department Of Computer Sc & Engg. Department Of Computer Sc &Engg. Kurukshetra Universty, India. March -2013.

19. Sang-Jun, H. and C. Sung-Bae, "Evolutionary neural networks for anomaly detection based on the behavior of a program", IEEE Trans. Syst. Man Cybernetics 36: 559-570. DOI: 10.1109/TSMCB.2005.860136, 2005.

20. Suguna, J. and A.M. Selvi, "Ensemble fuzzy clustering for mixed numeric and categorical data", Int. J. Comput. Appli., 42: 19-23. DOI: 10.5120/5673-7705, 2012.

21. Vladimir, N.V, "The Nature of Statistical Learning Theory", 2nd Edn., Springer, New York, ISBN-10:0387945598, pp: 188, 1995.

22. Wang, G., J. Hao, J. Ma and L. Huang, "A new approach to intrusion detection using Artificial Neural networks and fuzzy clustering", ExpertSyst.Appli.,37:62256232. DOI:10.1016/j.eswa.2010.02.1- 0 2, 2010.

23. Xindong, W., V. Kumar and J.R. Quinlan, "Top 10 algorithms in data mining" Knowl. Inform. Syst., 14: 1-37. DOI: 10.1007/s10115-007-0114-2, 2008.