International Journal of Advanced Computer Technology (IJACT)

ISSN:2319-7900

Intrusion Detection Using Data Mining Techniques

Krishna Kant Tiwari [1], Susheel Tiwari [2], Sriram Yadav [3]
[1] Student of Millennium Institute of Technology, RGPV University, Bhopal
[2] Asst. Professor (CSE), Millennium institute of technology, RGPV University, Bhopal
[3] Asst. Professor & Head (CSE), Millennium institute of technology, RGPV University, Bhopal

Abstract

In these days an increasing number of public and commercial services are used through the Internet, so that security of information becomes more important issue in the society information Intrusion Detection System (IDS) used against attacks for protected to the Computer networks. On another way, some data mining techniques also contribute to intrusion detection. Some data mining techniques used for intrusion detection can be classified into two classes: misuse intrusion detection and anomaly intrusion detection. Misuse always refers to known attacks and harmful activities that exploit the known sensitivity of the system. Anomaly generally means a generally activity that is able to indicate an intrusion. In this paper, comparison made between 23 related papers of using data mining techniques for intrusion detection. Our work provide an overview on data mining and soft computing techniques such as Artificial Neural Network (ANN), Support Vector Machine (SVM) and Multivariate Adaptive Regression Spline (MARS), etc. In this paper comparison shown between IDS data mining techniques and tuples used for intrusion detection. In those 23 related papers, 7 research papers use ANN and 4 ones use SVM, because of ANN and SVM are more reliable than other models and structures. In addition, 8 researches use the DARPA1998 tuples and 13 researches use the KDDCup1999, because the standard tuples are much more credible than others. There is no best intrusion detection model in present time. However, future research directions for intrusion detection should be explored in this paper.

Keywords— **intrusion detection, data mining, ANN**

1. INTRODUCTION

With the rapid expansion of the computer network during the past few years, the information security issue becomes more and more important. There are many research topics for network security. Like as, data encryption, vulnerability database, intrusion detection, etc. Intrusion detection is one of the major information security problems. IDS (Intrusion Detection System) assist the system in resisting external attacks. Existing IDS can be divided into two categories according to the detection approaches: anomaly detection and misuse detection or signature detection. [1][23]

Data mining techniques can be used for misuse and anomaly intrusion detection. Misuse refers to known attacks and harmful activities that exploit the known sensitivities of the system. In misuse detection, each instance in a data set is labeled as “normal” or” intrusion” and a learning algorithm is trained over the labeled data. (Dokas et al.). Anomaly means a usual activity in general that could indicate an intrusion. An advantage of misuse detection techniques is their high degree of accuracy in detecting known attacks and their variation. [4][8]

As there are many number of ID techniques using data mining techniques, the unknown technique and system could be thought of as a baseline for future prospect. As a result, the purpose of this paper is to review related papers of using data mining for intrusion detection. The contribution of this research paper is to provide a comparison of IDS in terms of data mining IDS techniques used for future research directions. This paper is organized as follows. Section 2 overviews the data mining techniques for intrusion detection. Section 3 compares related work of IDS. Section 4 discusses about the comparative results in the section and the conclusion is also provided.

2. RELATED WORK

2.1 IDS Architecture

An Intrusion Detection System is used to detect all types of malicious network traffic and computer usage that cannot be detected by an ordinary firewall. It includes network attacks against sensitive services, data driven attacks on computer applications, host based attacks such as privilege and permissions escalation, unauthorized logins and access to sensitive files, and malware (viruses, Trojan horses, and worms). IDS are the best fine grain filter placed inside the protected computer network, looking for known or powerful threats in network traffic and/or audit data recorded by hosts. The IDS architecture is shown in below.

2. 2 Data mining used in IDS

Much number of data mining techniques can be used in intrusion detection, each with its own specific advantage. The following lists some of the techniques and the motives for which they may be employed.

Classification: Creates a classification of tuples. It could be used to detect individual attacks, but as described by previous sample experiments in the literature indication it is produce a high false alarm rate. This problem may be reduced by applying fine-tuning techniques such as boosting.

Association: Describes relationships within tuples. Detection of irregularities may occur when many tuples exhibit previously unseen relationships. Grouping: Groups tuples that exhibit similar properties according to pre-described metrics. It can be used for general analysis similar to categorization, or for detecting outliers that may or may not represent attacks. Figure.

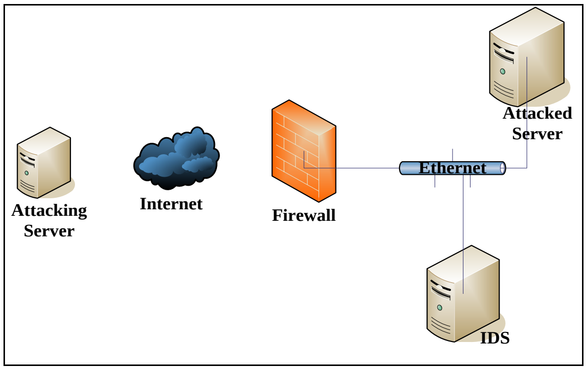

Figure1, IDS architecture

.

2.3 Models

2.3.1 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is relatively crude electronic models based on the neural structure of the brain. The brain basically learns from his experience. This is natural proof that some problems that are beyond the scope and range of current computers are indeed solvable by small energy efficient packages. This brain modeling a technical way to develop machine solutions. This new arrival approach to computing also provides a more graceful degradation during system overload than its more habitual counterparts.

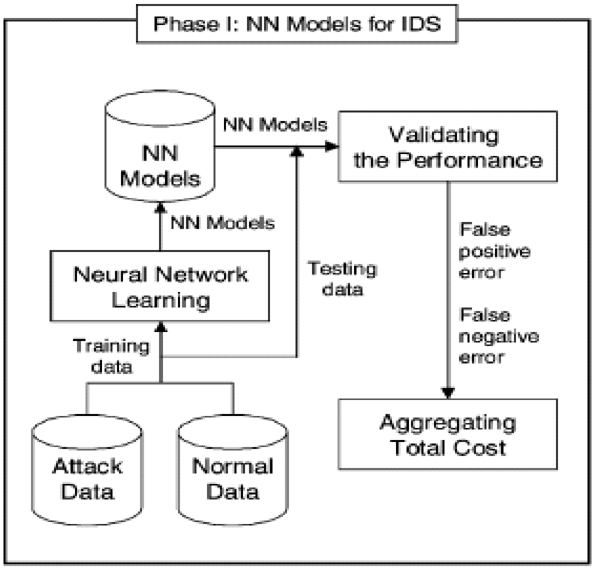

Figure 2, Shows a neural network model used for IDS. [11]

A neural network is an interrelated group of artificial neurons that uses a mathematical model or computational model for information processing based on a connection approach to computation. A neural network could not contains domain knowledge in the beginning, but it can be supervised to make decisions by mapping example pairs of input data into example output vectors, and estimating its weights so that it maps each input example vector into the corresponding output example vector approx. (Hecht-Nielsen, 1988).

2.3.2 Support vector machines (SVM)

The SVM approach converts data into a feature space F that usually has a large dimension. This is interesting to note that SVM generalization depends on the geometrical properties of the supervised data, not on the dimensions of the input space [10]. Detail information of SVM can be found in [26].

2.3.3 Multivariate adaptive regression splines (MARS)

Splines can be considered as an innovative mathematical process for complicated, hard curve drawings and function

ISSN:2319-7900

imprecise. The MARS model is a regression model using basic functions as predictors instead of the true copy data. The basic function convert makes it possible to selectively blank out certain regions of a variable by making them zero,

and allows MARS to focus on particular sub-regions of the record. It excels at finding optimal variable conversions and interactions, and the complicated data structure that often hides in high-dimensional data. [6]

.

3. REVIEW OF RELATED WORK

In this section, 3 tables create for compare related work of IDS, databases, including models and intrusion detection range criteria. Subsequently, our observations say data in the table and discuss it.

Existing IDS can be divided into two classes according to the detection approaches: anomaly detection and misuse detection. After reviewing these papers, almost all researches focus on anomaly detection. In the future, our research additionally focus on misuse detection

There are five main algorithms used in intrusion detection. Besides these five algorithms, one more category used, others, for other detection methods. Most researches in intrusion detection use ANNs Because, ANNs is much more reliable than other models & algorithms. Besides, the second most used model is SVMs. In the future, our research compares these models to recognize a“best” model for intrusion detection

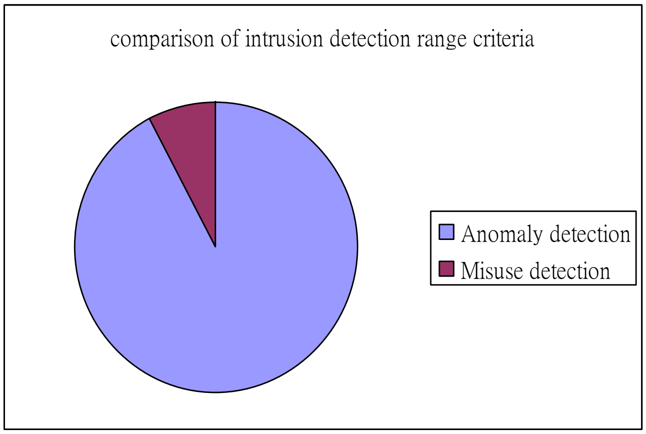

Figure 3, Comparison of intrusion detection range criteria

Most studies use the tuples of DARPA1998 and KDDCup1999, and others use some real data they collect by themselves. It is comparatively better to use a standard tuples than some personal data. Actually, MIT Lincoln Laboratory has already published DARPA2000. In the future, more experiments on intrusion detection made by using the DARPA2000 tuples.

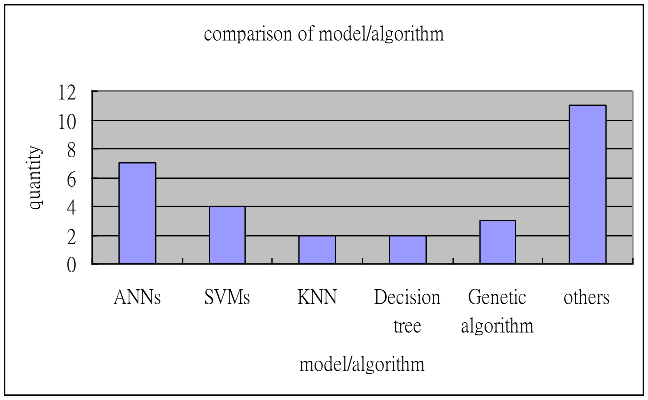

Figure 4, Comparison of models/algorithm.

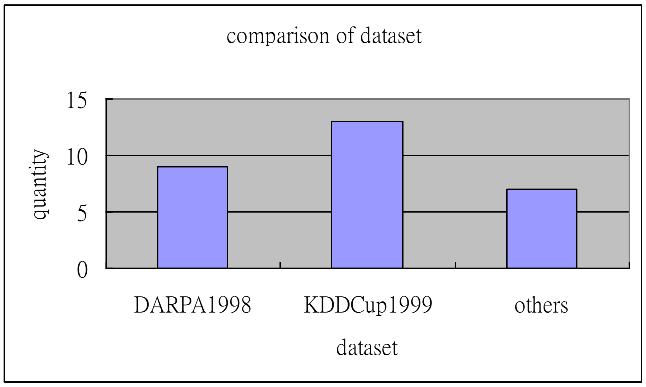

Figure 5, Comparison of used datasets.

4. CONCLUSIONS

The security of computer networks plays a planning role in modern computer system. Detection of intrusion attacks is the most important issue in computer network security. Existing IDS can be divided into two classifies according to the detection approaches: anomaly detection and misuse detection. There are several different methods to anomaly detection and misuse detection and misconfiguration. Approaches to anomaly detection have neural network, Statistics, Predictive pattern generation, and sequence matching and supervis-

ISSN:2319-7900

ing. In misuse detection, there are state transition analysis, pattern matching, model-based, keystroke monitoring and Expert system.

In this paper, comparison made in 23 papers for finding out the situation of intrusion detection now a day. After the comparison among these papers, observation shows most researches focus on anomaly detection, and use the tuples of DARPA1998 and KDDCup1999 mostly. In addition, most researches in intrusion detection use ANN. Because ANN is much more stable and reliable than other models and algorithms. Besides, the second most used model is SVM.

In the future, additionally focus on misuse detection. With the help of techniques are using in intrusion detection, compare these models in the paper to identify a “best” model for it. Besides, most researches use the tuples of DARPA1998 or KDDCup1999. Some experiments by using DARPA2000 or other tuples to increase objectivities later on. Probably, establish a new tuples to become the criterion of doing experiments in intrusion detection.

REFERENCES

[1] Anderson, J. "An introduction to neural networks." Cambridge: MIT Press. (1995)

[2] Chen, W.H., Hsu, S.H., and Shen, H.P. "Application of SVM and ANN for intrusion detection." Computer & Operations Research 32, 2617-2634. (2005)

[3] Cho, S.B., and Park, H.J. "Efficient anomaly detection by modeling privilege flows using hidden Markov model." Computer & Security 22, 44-55. (2003)

[4] Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., Srivastava, A.J., and Tan, P.N. "Data Mining for Network Intrusion Detection." Denning, D.E. "An Intrusion Detection Model, IEEE Trans-actions on Software Engineering." SE-13:222-232, 1987.

[5] Depren, O., Topallar, M., Anarim, E., Ciliz, M.K. "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks." Expert System with Applications 29, 713-222.

[6] Didaci,L.,Giacinto,G.and Roli, F. "Ensemble Learning for Intrusion Detection in Computer Networks"(2002), Proc. of AIIA, Siena, Italy, Friedman, J.H. "Multivariate adaptive regression spines." Anal Stat, 19:1–141. (1991)

[7] Han, S.J., and Cho, S.B. "Detecting intrusion with rule-based integration of multiple models." Computer & Security 22,613-623. (2003)

[8] Javitz, H.S.and Valdes, A.The NIDES Statistical Compo-nent: Description and Justification, Technical Report, Computer Science Laboratory, SRI International. (1993)

[9] Jiang, S.Y., Song, X., Wang, H., Han, J.J., and Li, Q.H. "A clustering-based method for unsupervised intrusion" Pattern Recognition Letters 27,802-810. (2006)

[10] Joachims, T. "SVM light is an implementation of support vector machines (SVMs) in C". University of Dortmund" Collaborative Research Center on Complexity Reduction in Multivariate Data (SFB475), http://ais.gmd.de/,thorsten/svm_light 2000.

[11] Joo, D., Hong, T., and Han, I. "The neural network models for IDS based on the asymmetric costs of false negative errors and false positive errors." Expert Systems with Applications 25, 69–75 2000.

[12] Lee, W., and Stolfo, S.J... "Data Mining Approaches for Intrusion Detection." Seventh USENIX Security Symposium (SECURITY ‘98), San Antonio, TX. (1998)

[13] Li, X.B. "A scalable decision tree system and its application in pattern recognition and intrusion detection." Decision Support System 41,112-130. (2005)

[14] Lippmann, R.P., and Cunningham, R.K. "Improving Intrusion Detection Performance Using Keyword Selection and Neural Network." Computer Network 34,597-603. (2000)

[15] Lippmann, R.P., Haines, J.W., Fried, D.J., Korba, J., and Das, K. "The 1999 DARPA off-line intrusion detection evaluation." Computer Networks 34,579-595. (2000)

[16] Liao, Y., and Vemuri, V.R. "Use of K-Nearest Neighbor classifier for intrusion detection." Computer & Security 21,439-448. (2002)

[17] Liu, Y., Chen, K., Liao, X. and Zhang, W. "A genetic clustering method for intrusion detection" Pattern Recognition 37,927-942. (2004)

[18] Mukkamala, S., Sung, A.H., and Abraham, A. "Intrusion detection using an ensemble of intelligent paradigms." Computer Applications 28,167-182. (2005)

[19] NikulinV"Threshould-based clustering with merging and regularization in application to network intrusion detection." Computational Statistics & Data Analysis. (2005)

[20] Ozyer, T., Alhajj, R., and Barker, K. "Intrusion detection by intelligent boostering genetic fuzzy classfier and data mining criteria for rule pre-processing." Journal of Network and Computer Applications. (2005)

[21] Perdisci, R., Giorgio, G., and Roli, F. "Alarm clustering for intrusion detection systems in computer net-

works." Engeering Applications of Artificial Intelligent 19,429-438. (2006)

[22] Pietraszek, T., and Tanner, A. "Data mining and machine learning- Toward reducing false positives in intrusion detection." Information Security Technical Report 10,169-183. (2005)

[23] Rhodes, B., Mahaffey, J., and Cannady, J. "Multiple self-organizing maps for intrusion detection Proceed ings of the 23rd national information systems security conference" Baltimore, MD, 2005.

[24] Ryan, J., Lin, M.J., and Mikkulainen, R."Intrusion Detection with Neural Network" 2005.

[25] Thompson, H.H., Whittaker, J.A.,Andrews, M. "Intrusion detection: Perspectives on the insider threat." Computer Fraud & Security, 2004, pp13-15. (2004)

[26] Vaprik, V. "Statistics learning theory." John Wiley, New York. (1998)

[27] Zhang, C. Jiang, J., and Kamel, M. "Intrusion detection using hierarchical networks." Pattern Recognition Letter 26,779-791, 2005.

[28] Zhang, Z., and Shen, H. "Application of online-training SVMs for real-time intrusion detection with different considerations." Computer Communications 28, 1428-1442. (2005)

Biographies

1 KRISHNA KANT TIWARI received the B.E. degree in information technology from the University of RGPV, Bhopal, M.P, in 2010. He is a scholar of Computer Science & Engineering at University of RGPV, Bhopal, M.P. India. His research areas include Intrusion Detection System.

3 SRIRAM YADAV received the B.E. degree in Computer Science Engineering from Gandhi institute of engineering and technology, gunapur, dist-Rayagada, Orissa in 2006. He is received the M.tech Degree in Computer Science Engineering from the P.G. Dept. of Computer Science, B.U. Ber hampur Orissa in 2009. He is presently pursuing Phd in Computer Science Engineering from pacific University of Higher Education and Research, Udaipur Rajsthan, India.

2 SUSHEEL TIWARI received the M.Tech degree in information technology from the U.I.T., B.U., Bhopal, M.P, in 2010, presently pursuing Phd Degree in Computer Science & engineering from Mewar University Chittorgarh (Rajasthan) India