A Review: AIS Based Intrusion Detection System

Sandeep Singh, Jasvinder Pal Singh, Gaurav Shrivastva

RKDF Institute of Science and Technology
Bhopal

Sandeep_rkdf@yahoo.com

Abstract

Prevention of security breaches completely using the existing security technologies is unrealistic. As a result, intrusion detection is an important component in network security. However, many current intrusion detection systems (IDSs) are signature-based systems, The signature based IDS also known as misuse detection looks for a specific signature to match, signalling an intrusion. Provided with the signatures or patterns, they can detect many or all known attack patterns, but they are of little use for as yet unknown attack methods. The rate of false positives is small to nil but these types of systems are poor at detecting new attacks, variations of known attacks or attacks that can be masked as normal behaviour. In this paper we evaluate the performance of various network based IDS technique and give an bird eye over existing IDS technique and their terminology.

Keyword: IDS, Artificial Immune System, Navie Bayes.

I. Introduction

Intrusion detection is defined [1] as the process of intelligently monitoring the events occurring in a computer system or network, analysing them for signs of violations of the security policy. The primary aim of Intrusion Detection Systems (IDS) is to protect the availability, confidentiality and integrity of critical networked information systems.

A. Types of Intrusion Detection Systems

- Network-based Intrusion Detection Systems (NIDSs)
- Host-based Intrusion Detection Systems (HIDSs).

These systems can be classified based on which events they monitor, how they collect information and how they deduce from the information that an intrusion has occurred. IDSs that scrutinize data circulating on the network are called Network IDSs (NIDSs), while IDSs that reside on the host and collect logs of operating system-related events are called Host IDSs (HIDSs).

II. Intrusion Detection Techniques

All intrusion detection systems are one of the two detection techniques.

- Signature based IDS
- Statistical anomaly based IDS

Signature based IDS: The signature based IDS also known as misuse detection looks for a specific signature to match, signalling an intrusion. Provided with the signatures or patterns, they can detect many or all known attack patterns, but they are of little use for as yet unknown attack methods. Most popular intrusion detection systems fall into this category. This means that an IDS using misuse detection will only detect known attacks or attacks that are similar enough to a known attack to match its signature.

Statistical anomaly based IDS: Another approach to intrusion detection is called anomaly detection. Anomaly detection applied to intrusion detection and computer security has been an active area of research since it was originally proposed by Denning (1987). Anomaly detection algorithms have the advantage that they can detect new types of intrusions as deviations from normal usage. In this problem, given a set of normal data to train from, and given a new piece of test data, the goal of the intrusion detection algorithm is to determine whether the test data belong to ‘‘normal’’ or to an anomalous behaviour. However, anomaly detection schemes suffer from a high rate of false alarms.

This occurs primarily because previously unseen (yet legitimate) system behaviour are also recognized as anomalies, and hence flagged as potential intrusions.

III. Evaluation Metrics

Metrics mainly used to evaluate the performance of classifier:

- The true positives (TP) and true negatives (TN) are correct classifications.

- A false positive (FP) occurs when the outcome is incorrectly predicted as yes (or positive) when it is actually no (negative).
- A false negative (FN) occurs when the outcome is incorrectly predicted as negative when it is actually positive.
- Recall: The percentage of the total relevant documents in a database retrieved by your search. If you knew that there were 1000 relevant documents in a database and your search retrieved 100 of these relevant documents, your recall would be 10%.

$$\text{Recall} = TP/(TP+FN)$$

- Precision: The percentage of relevant documents in relation to the number of documents retrieved. If your search retrieves 100 documents and 20 of these are relevant, your precision is 20%.

$$\text{Precision} = TP/(TP+FP)$$

- F-measure: The harmonic mean of precision and recall

$$f = \frac{2 * Recall * precision}{Recall + Precision}$$

- The true positive rate is TP divided by the total number of positives, which is TP + FN.
- The false positive rate is FP divided by the total number of negatives, FP + TN.
- ROC area In ROC analysis we plot true positive ratio (TPR) against, false positive ratio (FPR).

- The overall success rate is the number of correct classifications divided by the total number of classifications:

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

- In a multiclass prediction, the result on a test set is often displayed as a two dimensional confusion matrix with a row and column for each class. Each matrix element shows the number of test examples for which the actual class is the row and the predicted class is the column. Good results correspond to large numbers down the main diagonal and small, ideally zero, off-diagonal elements.

		yes	no
Actual Class	yes	True Positive	False Negative
	no	False Positive	True Negative

Figure1: Predicted class

A. Cost Matrix

All misclassifications had equal weights, target values or class labels that appear less frequently would not be privileged. You might obtain a model that misclassifies these less frequent target values while achieving a very low overall error rate. To improve classification decision trees and to get better models we generate an appropriate cost matrix to balance the distribution of class labels when a decision tree is trained. You can also manually adjust the cost matrix. A cost matrix or error matrix is also useful when specific classification errors are more severe than others. The Classification mining function tries to avoid classification errors with a high error weight. The trade-off of avoiding 'expensive' classification errors is an

increased number of 'cheap' classification errors. Thus, the number of errors increases while the cost of the errors decreases in comparison with the same classification without a cost matrix. [6]

IV. Related Work Done

Application and development of specialized machine learning techniques is gaining increasing attention in the intrusion detection community. A variety of learning techniques proposed for different intrusion detection problems can be roughly classified into two broad categories:

- supervised (classification)
- Unsupervised (anomaly detection and clustering).

The supervised learning methods significantly outperform the unsupervised ones if the test data contains no unknown attacks. Following are the data mining related work for intrusion detection.

A. Agent based AIS for IDS

Chung-Ming Ou , and Yao-Tien Wang, et. al [8] proposed Agent-based artificial immune system (ABAIS) to apply over intrusion detection systems (IDS) having Three agents, namely, Ag agent, DC agent and TC agents are coordinated to exchange information of intrusion detections. The intelligence behind such system is based on the danger theory of human immune systems. In particular computations of danger values with dynamic thresholds will reduce the false positive rate of danger signals issued by computer hosts.

B. Artificial Anomalies

In [9], the authors have proposed an algorithm to generate artificial anomalies to force the inductive learner to find out a more accurate boundary between known classes (normal connections and known intrusions) and anomalies. Their experiment on the KDD99 data set shows that the model is capable of detecting more than 77% of all unknown intrusion classes with over 50% accuracy per intrusion class. However, the way to generate anomalies is not clear.

C. ***Adapted IDS Model***

By Lei Deng De-yuanGao [10] proposed Immune based Adaptive IDS Model (IAIDSM) is using Enhanced Fast Adaptive Clustering Algorithm and Algorithm of Mining Fuzzy Associate. The Immune based Adaptive IDS Model (IAIDSM)would be accurate, low in false alarms, not easily cheated by small variations in patterns, adaptive and be of real time. Analyzing the training data obtaining from internet, the self-behavior set and non-self-behavior set can be obtained by the partitioned clustering algorithm, then it extracts Self and non-self-pattern sets from these two behavior sets by association rules and sequential patterns mining.

D. Network Intrusion Detection Using Naive Bayes

In [11], the authors have proposed a framework of NIDS based on Naive Bayes algorithm. The framework builds the patterns of the network services over data sets labelled by the services. With the built patterns, the framework detects attacks in the datasets using the naive Bayes Classifier algorithm. Authors carry out their experiment on 10% of the KDD'99 dataset, which contains 65,525 connections. For their experiments, they choose the naïve Bayes Classifier in WEKA Compared to the neural network based approach, our approach achieve higher detection rate, less time consuming and has low cost factor. However, it generates somewhat more false positives. As a naive Bayesian network is a restricted network that has only two layers and assumes complete independence between the information nodes. This poses a limitation to this research work.

E. Analysing Information Security Issues Using Data Mining Techniques

In [12] authors present an example of data mining techniques application in the framework of network regarding vulnerabilities. The software used is Weka, a collection of machine learning algorithms for data mining tasks implemented in Java language. The input data are taken from the NVD – National Vulnerability Database. The Common Vulnerability Scoring System (CVSS) provides an open framework for communicating the characteristics and impacts of IT

vulnerabilities. Its quantitative model ensures repeatable accurate measurement while enabling users to see the underlying vulnerability characteristics that were used to generate the scores. Thus, CVSS is well suited as a standard measurement system for industries, organizations, and governments that need accurate and consistent vulnerability impact scores. Two common uses of CVSS are prioritization of vulnerability remediation activities and in calculating the severity of vulnerabilities discovered on one's systems. The National Vulnerability Database (NVD) provides CVSS scores for almost all known vulnerabilities. The variables used in the statistics are the base score, the exploit sub score, the impact sub score, the product name (operating system and other products) and severity 219 (gravity of the attack). The authors intend to determine how the gravity of the attacks depends on the others four variables (base score, exploit score, impact sub score and product name).The database used in this application is in csv format and contains five attributes in different data format (nominal, numeric) and 11000 records.

Intrusion detection is an important task for information infrastructure security. One major challenge in intrusion detection is that we have to identify and classify the hidden intrusions from a huge amount of normal communication activities. The data mining techniques are viable solutions for determine the severity of the attacks and they can be included in IDS, generating the development of IDS based on Data Mining.

F. Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context

A simulation study was performed in [13] to assess the performance of a comprehensive set of machine learning algorithms on the KDD 1999 Cup intrusion detection dataset. Simulation results demonstrated that for a given attack category certain classifier algorithms performed better. Consequently, a multi-classifier model that was built using most promising classifiers for a given attack category was evaluated for probing, denial-of-service, user-to-root, and remote-to-local attack categories. The proposed multi-expert

classifier showed improvement in detection and false alarm rates for all attack categories as compared to the KDD 1999 Cup winner. Furthermore, reduction in cost per test example was also achieved using the multi-classifier model. However, none of the machine learning classifier algorithms evaluated was able to perform detection of user-to-root and remote-to-local attack categories significantly (no more than 30% detection for U2R and 10% for remote-to-local category). In conclusion, it is reasonable to assert that machine learning algorithms employed as classifiers for the KDD 1999 Cup data set do not offer much promise for detecting U2R and R2L attacks within the misuse detection context.

G. Adaptive Framework for Network Intrusion Detection by Using Genetic-Based Machine Learning Algorithm

In [14], a new algorithm was introduced to detect network intrusions and was successfully demonstrated on KDD 99 Dataset, training and testing data. Genetic-based machine learning algorithm (GBML) which offers a good ability to be adapted to changing environments, robustness to noise and ability to identify unknown attacks. The objective of this paper was to incorporate different techniques into classifier system to detect and classify intrusion from normal network packet. Among several techniques, steady state genetic-based machine leaning algorithm (SSGBML) which will be used to detect intrusions. Steady State Genetic Algorithm (SSGA) and Zeroth Level Classifier system (ZCS) are investigated. SSGA is used as a discovery mechanism for classifiers, while ZCS plays the role of detector by matching incoming environment message with classifiers to determine whether it is normal or intrusion Also, matching difference between environment message and classifier rules became adaptive according to DR values. Discover engine has been improved by using SSGA instead of SGA taking into account the suitable method for selection. Also, when performing some training on SGA, we dramatically reached premature convergence early. So, training phase was stopped

and not continued. The proposed work focused on reducing the number of features to be used in classifying and detecting various attacks types.

H. A Hybrid Network Intrusion Detection Technique Using Random Forests

The author has proposed framework of the hybrid detection is shown in following Figure 2. First, observed activities are fed to the misuse detection component. The component applies the random forests algorithm to detect known attacks by matching the patterns of attacks. Other items (uncertain items) that do not match any pattern are fed to the anomaly detection component to detect unknown intrusions using the outlier detection.

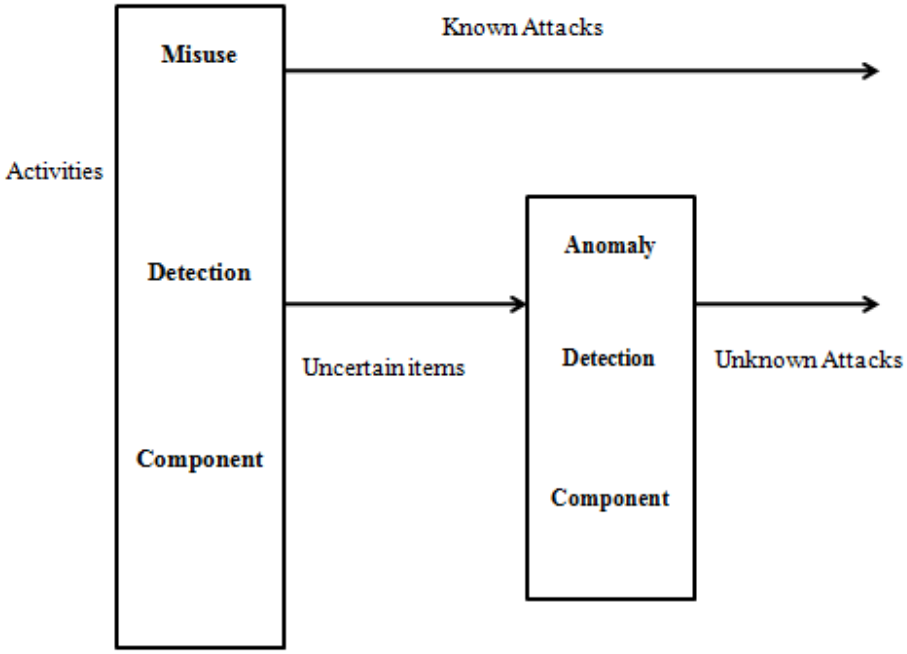

Figure 2: Hybrid classifier

V. Proposed methodology

To improve the performance of misuse detection, [15] employed the feature selection algorithm to calculate the value of variable importance over the training set. The experimental results show that the proposed hybrid approach can achieve high detection rate with low false positive rate, and can detect novel intrusions. However, some intrusions that are very similar with each other cannot be detected by the anomaly detection. That is a limitation of the outlier detection provided by the

random forests algorithm. To solve the above problem, they suggested that other data mining algorithms such as clustering algorithm could be investigated in the future.

IDS focus on exploiting attacks, or attempted attacks, on networks and systems, in order to take effective measures based on the system security policies, if abnormal patterns or unauthorized access is being suspected. A lot of methods and techniques have been proposed for the effective designing of IDS. But all technique suffered common problem that problem is detection and prediction of false positive and false negative rate is high. Due to this problem the given methodologies are not used in generalize form. So we modified one of the existing second generation AIS algorithm called Dendritic Cell Algorithm for controlling a generation of false alarm generation and also improve classification rate of data more accurately. The Dendritic Cell Algorithm categories efficiently into the normal and abnormal data and Dempster-Belief theory is used to compute the probability of evidences that indicate support the attack or normal class. Along with that proposed methodology encapsulate SVL Classifier to classify normal class and attack class based on D-S theorem. The use of Dempster Belief theory steadily spreads out, mostly because it is used to cope with large amounts of uncertainties that are inherent of continuously changing environment.

VI. Conclusion

Increased connectivity and the use of the internet have exposed the organization to subversion, there by necessitating the use of intrusion detection system to protect information system and communication network from malicious attacks and unauthorized access. An intrusion detection system (IDS) is a security system that monitors computer systems and network traffic, analyses that traffic to identify possible security breaches, and raise alerts. An IDS triggers thousands of alerts per day making it difficult for human users to analyses them and take appropriate actions. It is therefore important to reduce the false alarm alerts, intelligently integrate and correlate them and to present a high level view of the detected security issue to the administrator.

References

[1] Li Rui, Luo Wanbo, "Intrusion Response Model based on AIS" in International Forum on Information Technology and Applications, IEEE , 2010.

[2] YUAN Hui, LIU Jian-yong, "Intrusion Detection Based on Immune Dynamical Matching Algorithm" in International Conference on E-Business and E-Government, IEEE, 2010.

[3] Li Zhi-tang , Li Yao, Wang Li , "A Novel Fuzzy Anomaly Detection Algorithm Based on Artificial Immune System", IEEE, 2005.

[4] Baoyi WANG,Shaomin ZHANG, "A New Intrusion Detection Method Based on Artificial Immune System" in IFIP International Conference on Network and Parallel Computing – Workshops, IEEE, 2007.

[5] Yu-fang zhang, gui-hua sun, zhong-yang xiong, "A novel method of intrusion detection based on artificial Immune system" in proceedings of the fifth international conference on machine learning and cybernetics, dalian, IEEE, 2006.

[6] Guo Chen, Peng Shuo, Jiang Rong & Luo Chao, "An anomaly detection system based on dendritic cell algorithm" in Third International Conference on Genetic and Evolutionary Computing, IEEE, 2009.

[7] Real-time feature selection in traffic classification, ZHAO Jing-jing1, HUANG Xiao-hong2, SUN Qiong2, MA Yan1, 2, 3 The Journal of China, Universities of Posts and Telecommunications, Vol 15, Supplement 1 2008, pp. 68-72.

[8] Chung-Ming Ou, Yao-Tien Wang C.R. Ou , "Intrusion Detection Systems Adapted from Agent-based Artificial Immune Systems", IEEE International Conference on Fuzzy Systems, 2011, pp 115 -122.

[9] Wei Fan, Matthew Miller, Salvatore J. Stolfo, Wenke Lee, and Philip K. Chan. Using artificial anomalies to detect unknown and known network intrusions. Proceedings of the IEEE International Conference on Data Mining, 2001.

[10] Lei Deng, De-yuan Gao, "Research on Immune based Adaptive Intrusion Detection System Model", International Conference on Networks Security, Wireless Communications and Trusted Computing, 2009, pp-488-492

[11] Mrutyunjaya Panda1 and Manas Ranjan Patra2. Network Intrusion Detection Using Naive Bayes. IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.

[12] Daniela Şchiopu1, Irina Tudorl. Analyzing Information Security Issues Using Data Mining Techniques. The 3rd International Conference on Virtual Learning, ICVL, 2008.

[13] Mahesh kumar Sabhnani. Gursel Serpen Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context. Vol 8 , no. 4, 2004

[14] Wafa' S. Al-Sharafat, Reyadh Sh.Naoum. Adaptive Framework for Network Intrusion Detection by Using Genetic-Based Machine Learning Algorithm. IJCSNS International Journal of Computer Science and Network Security, VOL.9, no.4, April 2009.

[15] Jiong Zhang and Mohammad Zulkernine A Hybrid Network Intrusion Detection Technique Using Random Forests. Proceedings of the First International Conference on Availability, Reliability and Security, IEEE, 2006.