(Print)

JUCIT Vol. 8(2), 16-19 (2017). Periodicity-2-Monthly

(Online)

JOURNAL OF ULTRA COMPUTER & INFORMATION TECHNOLOGY

An International Open Free Access Peer Reviewed Research Journal of Computer
Science Engineering & Information Technology

website:- www.compitjournal.org

Estd. 2010

A Survey of Supervised Machine Learning's Algorithms for Intrusion Detection

AMRIN MANSOORI¹, ANKITA HUNDET², BABITA PATHIK³ and SHIV KUMAR⁴

¹M. Tech Scholar, Department of Computer Science & Engineering, Lakshmi Narain College of Technology & Excellence Bhopal (M.P), India

^{2,3,4}Faculty of Computer Science & Engineering, Lakshmi Narain College of Technology & Excellence Bhopal (M.P), India

Email of Corresponding Author :- amrinmansoori7@gmail.com

<http://dx.doi.org/10.22147/jucit/080201>

Acceptance Date 11th April, 2017,

Online Publication Date 15th April, 2017

Abstract

As network attacks have increased in number and severity over the past few years, intrusion detection system (IDS) is increasingly becoming a critical component to secure the network. Due to large volumes of security audit data as well as complex and dynamic properties of intrusion behaviours, optimizing performance of IDS becomes an important open problem that is receiving more and more attention from the research community. This paper compares the performance of Intrusion Detection System (IDS) Classifiers using various feature reduction techniques. To enhance 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. These include: classifying and clustering Algorithms Naïve Bayes, Simple k mean, Decision tree and J48, Linear Discriminate Analysis, and Independent Component Analysis. Many Intrusion Detection Systems are based on neural networks. However, they are computationally very demanding. 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. We try to build a system which creates clusters from its input data by labelling clusters as normal or anomalous data instances and finally used these clusters to classify unseen network data instances as either normal or anomalous¹. Both training and testing was done using different subsets of KDD Cup 99² data which is very popular and widely used intrusion attack dataset.

Key word: Supervised Machine Learning, Unsupervised Machine Learning, Network Intrusion Detection, Network Security.

1. Introduction

Intrusions can be defined as actions that attempt to bypass the security mechanisms of computer systems. Intrusions may take many forms: attackers accessing a system through the Internet or insider attackers; authorized

(official) users attempting to gain and misuse non-authorized privileges. So, we say that intrusions are any set of actions that threaten the integrity, availability, or confidentiality of a network resource. Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions. Intrusion

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-sa/4.0>)

detection systems (IDS) raise the alarm when possible intrusions occur¹⁻³.

He concludes that the “evaluation of human integration” is necessary to reduce classification errors. His experimental results showed that RBF-SOM achieves similar or even better results, compared to just an RBF. Network Security maintenance is one of the major safety concerns for neutralizing any unwanted activities. It is not only for protecting data and network privacy issues but also for avoiding any hazardous situations. From January through June 2010 Microsoft security intelligence report shows that the infection trends are still increasing on average around the world at a higher rate⁶. For decades, Network security is one of the major issues and different types of developed systems are being implemented. Network intrusion is an unauthorized activity over the network that steals any important and classified data. Also sometimes it's the reason of unavailability of network services. The unexpected anomaly occurs frequently and a great loss to internet cyber world in terms of data security, the safety of potential information's etc. There are several types of method proposed for network intrusion detection. The anomaly network intrusion detection is a major part of network security^{7, 8}.

A lot of research into artificial neural networks (ANNs) has been undertaken. In⁴, artificial neural networks and support vector machine (SVM) algorithms were applied to intrusion

Detection (ID), using a frequency-based encoding method, on the DARPA dataset. The authors use 250 attacks and 41,426 normal sessions and the detection rate (DR) varied from 100% to 43.6% with the false positive rate (FPR) ranging from 8.53% to 0.27% under different settings. In⁵, the author concludes that the combination of a radial basis function (RBF) and self organizing map (SOM) is useful as an intrusion detection model.

Figure 1: A linear Support Vector Machine⁸

2. Literature Survey :

2.1 Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection :

In this paper Author conclude that Intrusion detection model is a predictive model used to predict the

network data traffic as normal or intrusion. In this algorithms (Machine Learning) are used to build accurate models for clustering, classification and prediction. In this paper Author used number of algorithms namely Logistic Regression, Gaussian Naive Bayes, Support Vector Machine and Random Forest. These algorithms are tested with NSL-KDD data set. The standard intrusion detection data set KDDCUP9923 has redundant records. This may lead to unfair result of the machine learning algorithms. So the supervised machine learning algorithms are tested NSL-KDD24 data-set which is the advanced version of the KDDCUP99 intrusion detection data-set. Supervised machine learning algorithms namely Logistic Regression, Gaussian Naive Bayes, Support Vector Machine and Random Forest are tested on NSL-KDD dataset, the new standard intrusion detection data-set. These algorithms are tested on Intel Core (TM) i5-3230M CPU @2.60 GHZ, 4GB RAM and coding is done by Python20.

2.2 Intrusion Detection Techniques in Wireless Sensor Network using Data Mining Algorithms: Comparative Evaluation Based on Attacks Detection⁹

Wireless sensor network (WSN) consists of sensor nodes. WSN suffers from several attacks, intrusion and security vulnerabilities. Intrusion detection system (IDS) is one of the essential security mechanisms against attacks in WSN. In this paper Author present a comparative evaluation of the most performing detection techniques in IDS for WSNs, the analyzes and comparisons of the approaches are represented technically. Attacks in WSN also are presented and classified into several criteria. To implement and measure the performance of detection techniques we prepare our dataset, based on KDD'99, into five steps, after normalizing our dataset, we determined normal class and 4 types of attacks, and used the most relevant attributes for the classification process.

3. Problem Identificationm:

A lot of research work has been done in the field. RBSs (Rule Based Systems) are well suited for event correlation to perform misuse detection. However, other techniques are better suited for anomaly detection, such as statistical methods and clustering. The ability to facilitate anomaly detection is one of the benefits that have motivated much research on machine learning for intrusion detection. In the last decade, an increasing amount of research on machine learning for misuse detection can also be observed in this review. The application of techniques such as Artificial Neural Networks (ANNs) to misuse detection offers some desirable flexibility in the detection process compared with conventional RBSs, i.e., variations of learned attacks can be detected. The in- flexibility of RBSs, due to operating with

"crisp" rules, has been considered one of their main drawbacks. However, this observation is no longer entirely accurate, since researchers have proposed several applications of fuzzy RBSs, which have also been shown to be capable of performing anomaly detection.

4. Proposed Methodology & Block Diagram :

Figure 2: Methodology & Block Diagram¹

In the above Block Diagram we have number of Level

- ❖ Input Data & Pass to different Algorithm for Process
- ❖ Apply Training for Different set of data
- ❖ Compare our results among different algorithms

In the above model we will discuss number of features and the attacks out of some we will concentrate more which is given below:

Denial of Service (DoS) attack: Over usage of the bandwidth or non availability of the system resources leads to the DoS attacks. Examples: Neptune, Teardrop and Smurf.

User to Root (U2R) Attack: Initially attacker access normal user account, later gain access to the root by exploiting the vulnerabilities of the system. Examples: Perl, Load Module and Eject attacks.

Probe Attack: Have an access to entire network information before introducing an attack. Examples: ipsweep, nmap attacks.

Root to Local (R2L) Attack: By exploiting some of the vulnerabilities of the network attacker gains local access by sending packets on a remote machine. Examples: imap, guess password and ftp-write attacks.

5. Conclusion

After reading number of research papers we conclude that, a supervised learning approach to the intrusion detection problem is investigated and demonstrated on the International Knowledge Discovery and Data Mining Tools Competition intrusion detection benchmark (the KDDCUP 99 dataset). To do so this we investigated two architectures; the first engine is a back-propagation neural network intrusion detection system (BPNNIDS) and the second is a RBFNNIDS. The two engines work under two basic data sets; one is limited to 19361 connections (records), which is the set we selected from the 10% version of the KDD Cup 1999 dataset whereas the other contains 494021 connections (records), which is the 10% version of the KDD Cup 1999 dataset. Algorithms that are tested with the NSL-KDD data-set. Effective classifier is identified by comparing the performances based on the precision, recall, F1-Score and accuracy. From the observed results it can be concluded that the Random forest classifier outperforms other classifiers for the considered data-set and parameters. It has the accuracy of 99%. The work can be extended by considering the classifiers for multiclass classification and considering only the important attributes for the intrusion detection.

References

1. Manjula C. Belavagi, Balachandra Muniyal "Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection" Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016).
2. D. Anderson, T. Frivold, and A. Valdes, Next-Generation Intrusion Detection Expert System (NIDES): a summary. SRI Technical Report No.: SRI-CSL-95-07, Menlo Park, CA: SRI International Computer Science Laboratory, (1995).
3. K. R. Kendall, "A database of computer attacks for the evaluation of intrusion detect systems," MS Thesis, Massachusetts Institute of Technology, Cambridge, MA, (1999).
4. W. H. Chen, S. H. Hsu, and H. P. Shen, "Application of SVM and ANN for intrusion detection," Computers and Operations Research, vol. 32, no. 10, pp. 2617-2634, (2005).
5. T. Horeis, "Intrusion detection with neural networks combination of self-organizing maps and radial basis

- function networks for human expert integration,” <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.191&rep=rep1&type=pdf>.
6. D. Batchelder, J. Blackbird, P. Henry, and G. MacDonald, “Microsoft Security Intelligence Report Volume 17,” Microsoft Secur. Intell. Rep., vol. 16, pp. 1–19, (2014).
7. G. Xiaoqing, G. Hebin, and C. Luyi, “Network intrusion detection method based on Agent and SVM,” 2010 2nd IEEE Int. Conf. Inf. Manag. Eng., pp. 399–402, (2010).
8. R. P. Lippmann and R. K. Cunningham, “Improving intrusion detection performance using keyword selection and neural networks,” Comput. Networks, vol. 34, pp. 597–603, (2000).
9. A. T. Yousef El Mourabit, Anouar Bouirden and N. E. Moussaidr, Intrusion Detection Techniques in Wireless Sensor Network Using Data Mining Algorithms: Comparative Evaluation Based on Attacks Detection, International Journal of Advanced Computer Science and Applications, vol. 6, no. 9, pp. 164–172, (2015).
10. G. Nadiammai and M. Hemalatha, Effective Approach Toward Intrusion Detection System Using Data Mining Techniques, Egyptian Informatics Journal, vol. 15, no. 1, pp. 37–50.