Survey on Feature Reduction Techniques of Intrusion Detection System

Rashmi Singh
Phd Scholar, Computer Science Department of LNCT University, Bhopal, MP, India

Dr. Praveen Kumar Mannepalli
Computer Science Department of LNCT University, Bhopal, MP, India

Abstract: ***As variety of service execution is done on digital networks, thus security of these plays an important role. Hence intrusion detection system were developed by researcher to extend the reliability of internet network users. This paper has focus on the network intrusion detection techniques proposed by researcher of this field. Different types of attacks are done by intruders on the system, network, cloud, etc. This paper finds some feature reduction survey likewise which will increase the intrusion detection accuracy. Classification of IDS techniques was conjointly performed. Numerous evaluation parameters were summarized for the comparison of methods.***

Index Terms- Anomaly Detection, ANN, Clustering, Genetic Algorithm, Intrusion Detection.

I. INTRODUCTION

Intrusion detection systems (IDSs) are one amongst the foremost important entities when it involves Information and Communications Technology (ICT) infrastructure protection against cyber attacks. IDSs weapon help defenders with fundamental means to detect offensive events and consequently trigger optimal counteraction plans against them [1], [2].

In fact, the everlasting battle between defenders and attackers has taken the shape of an ‘‘arm race’’, where each sides constantly upgrade their arsenals so as to prevail against one other. With the emergence of latest and new attacks ,academia and industry is continuously working to analyse and investigate new methodologies which are ready to closely monitor this race and adapt rapidly to the changes within the field.

In principle, IDSs constitutes two major categories, namely Anomaly Detection Systems and Misuse Detection Systems [3]. The previous regulate their detection engine to recognize as intrusive incidents who exhibit deviations from a predefined normal behavioral profile. This sort of IDSs are able to identify previously unseen attacks, but are known to produce high false alarm rates, rendering them a questionable solution especially for complex infrastructures, where the standardization of the normal profile is challenging. On the opposite hand, misuse IDSs depend on known signatures trying to designate traffic instances to legitimate or attack traffic classes. This kind of IDS lacks the capability of identifying new attack patterns or deviations from known ones, and their performance depends on the freshness of the signatures database.

Due to above reasons significant effort is required by the IDSs administrator to keep the misuse detection model up to date. If we additionally consider the very fact that the protected environment could be a dynamic ecosystem where new devices and/or services may appear or leave the network at any moment (e.g., the Internet of Things), it becomes clear that the adaptability issue becomes a burden on administrator’s shoulders [4]. This burden becomes even heavier because the growth of communication networks pushes IDSs into the big data era, where the increased volume of the transmitted data surpasses the bounds of human processing capabilities.

II. Types of Attacks

1. Insider attack

Sometimes people having the authorization to use the cloud service, though choose to gone through the insider way. This mainly done with the intention of using the unauthorized privileges and revealing the information to other clients or in market. An insider attack has been planned mostly by the employees of the competitors or the cloud administrator in the domain client company having right to access those. They also had hand in modifying the company’s information and documents.

The best-known example to clear about this insider attack is the Amazon Elastic Compute Cloud (EC2) (Slaviero, 2002) – an internal attack of DoS.

2. Cloud malware injection attack

In this attack, the attacker has the motive of not only accessing the information but also get control over it of the client data. For this attacker creates its own service implementation module for setting it into a client cloud system. For this uses SaaS/PaaS method or the virtual machine instance into the IaaS solution. To result in a performing malicious activity, attacker if gets succeed in his work of cloud fouling, the cloud will automatically accept and sends the hacker module information to the user. Due to which the begins of malicious activities performing by the attacker.

Types of attacks under this category:

Cross site scripting attack: XSS uses the HTML for the attack in which malicious code is injected into the data by using the Flash, JavaScript or others.

SQL injection attack: In this attack the attacker uses the input field of the database of the user. The most common example for such types of attacks are the attack occurred on the Sony play station in the year 2008 website.

Command injection attack: the name of this attack is given as per its role, because it injects the command and those commands are run according to the runtime environment or may create shell.

3. Abuse and Nefarious use of cloud services

The main difference in this attack than the insider is the attackers background, otherwise all is in common. In the insider attack the attacker is the authorised user of the data while in this attacker is the hacker which attacks the less

Published By:
International Journal of Engineering Research in Current Trends

secured database or poor clouds. As due to this no need of using expensive DoS and did brute forced attacks on the target.

4. Denial of service attack

This type of attacksis mainly done by the flooded networks having many packets like TCP, UDP, ICMP or their combinations.due to the risk of the intruder attack on the distributed services of the computer, some of them are not even available to the authorised users also. As this attack overloads all the systems, due to which legal users are unable to used them.These types of attacksprove very dangerous for the single cloud data and servers as many users depends on that cloud distributed network.

5. Side channel attack

This type of attack done with the cryptographic algorithm of the system. For this they used the special VMM service which is virtual machine manger which guides the user attack for the creation of virtualization layer. They placed a physical virtual machine on the targeted system, while VMM helps other users and supervises known as hypervisor.

6. User to root attack

In this attack, the attacker uses the sniffing password for the authentication of the targeted user's system. So, by combining traditional various methods for the raising of the privileges to the super user access acceptance. An example of such escalation technique is the smashing stack, in which a packet of the set-UID- root program that corrupts the address space, so that returning information from the instruction to subshell space.

7. A remote to Local attack

In this attacker takes the advantages of the targeted user local privileges. This attack is also known as remote to user attack. In this attacker sends packets to the user host and close the exposures of the access of asxlock, guest, xnsnoop, phf and sendmail.

III. RELATED WORK

Yogitha et. al. [9] presented interruption discovery framework with Support Vector Machine (SVM). Confirmation is ended by organizing surveys on NSL-KDD Cup'99 information gathering which is reformer type of KDD Cup'99 data index. By using this NSLKDD Cup'99 information gathering they have reduced spacious time essential to shape SVM exemplary by attainment appropriate pre-training on information gathering. In this organization SVM made bunching of information. By compulsion suitable part gathering attack location rate is opened up and false positive rate (FPT) is pointed. In this planned work writer has used Gaussian Circular Basis.

A.R. Jakhale, et. al [10] In this exertion the writer depicts a anamoly discovery framework and its two phases chiefly are training and testing. The slipping window and gathering is familiar to tending the web network move by pulling out the recurring examples using computations. The estimation is as authentic and used as a division of regular monitoring. The standard multi-design communicable computation has elevated location rate. At long last, boost the recognition rate and compact the fake aware rate.

Barolli et al [11] investigates the consumption of IDS using neural network for giving IDS arrangement in a Tor (The Onion Router) manage. Examinations did use a Tor server and client with back engendering NN to replicate exchanges over the Tor organizes while infectious for assessment. The structure planned is a ready ANN with data taken from Wireshark, at that position the server and client data are examined, and distinctions will identified an interruption or abuse. The conclusion from testing was productive in giving feasible accuracy when charged in the test situation.

Kaiyuan et. al. in [12] propose a network intrusion detection algorithm combined hybrid sampling with deep hierarchical network. Firstly, we use the one-side selection (OSS) to reduce the noise samples in majority category, and then increase the minority samples by Synthetic Minority Over-sampling Technique (SMOTE). In this way, a balanced dataset can be established to make the model fully learn the features of minority samples and greatly reduce the model training time. Secondly, we use convolution neural network (CNN) to extract spatial features and Bi-directional long short-term memory (BiLSTM) to extract temporal features, which forms a deep hierarchical network model.

Chuan Long [13] In this article, writer examine how to present an interruption recognition framework in light of thoughtful learning, and this exertion offer a thoughtful knowledge approach for interruption recognition using recurrent neural networks (RNN-IDS). In addition, this exertion inspect the execution of the model in balancing categorization and multiclass arrangement, and the amount of neurons and characteristic learning rate impacts on the implementation of the planned show. This effort compare it and those of J48, artificial neural network, arbitrary woodland, bolster vector machine, and further machine knowledge approach planned by history analysts on the standard information directory index.

IV. DATASET FEATURE REDUCTION TECHNIQUES

It refers to mapping of high-dimensional data to a lower dimensional space. [14]. Criterion for this may differ based on different dataset types or content.

Kaiyuan et al.in use a network intrusion detection algorithm in the paper, for this uses hierarchical network mixing with the hybrid sampling. For the balancing of the majority and minority sampling, first test with the OSS that is the one side selection and then apply SMOTE which is synthetic minority over sampling technique. This can be used to trained the features of minority sampling and model time decreasing that is OSS by reduce noise. After the sampling is done, the deep hierarchical network is applying for the special features that is CNN and BiLSTM. CNN (convolution neural network) and BiLSTM (Bi-directional long short-term memory) for the extracting 3D and temporal features.

Singular valuation deposition (SVD) is the procedure which is used to reduce the dimensionality in the data comes under the vector algebra by factorizing the matrix. The main motive of the SVD is the data analysis of the gene that is to identify and expel the mechanical constraints and narrate the significance. To calculate the covariance matrix eigen values and eigen vectors of the sample gene matrix SVD is applied. To maintain the higher variability of the

matrix the eigen values must be greater so as to reduce the eigen vector. Eigen vectors causes the PC that is prime principle components to reduce into smaller scales due to higher unpredictability of the vectors.

Partial Least Square Regression(PLSR) while researching there must be desirability of controlling and easy calculating of variables response, so as to easily find out other variables that are related. While doing so there must be MLR multiple layer regression technique used as due to less variable and collinearity of those, it's easy to maintain the responses of them. But when any of the above condition doesn't not satisfy, then MLR fails to used. Then researchers faced the problems and have to deal with the ill responses and many variables for predictive model. E.g. spectrograph

Linear Discriminant Analysis (LDA) is a method used majorly in subjects like Statistics, Machine Learning and where there is need of pattern recognition, dimensionality reduction as it distributes the classes of objects. It is also use for the linear classification.

Locally Linear Embedding (LLE) has the feature of converting the high dimensional data to low dimensional data with no loss of information. The most common method used for the dimension reduction is PCA (principle component analysis) in which data points are spanned in the data sets. By the great variance they are covered by the directions of the respective orthogonal projection of subspace low dimension of components or factors. For the nonlinear algorithm the stated methods are LLE and ISOMAP. These methods convert the high dimensional data into low dimensional subspace by placing the important components into the nonlinear way. LLE can be coordinated the data after the converting into the smooth manifold and into individual local co-ordinates.

Generic algorithm research on the Darwinian evolution and natural selection has stated many theories and models for deciding optimization. This algorithm is the substitute for those problem on the population of the mutation, selection and recombination applications. They are applied for the pattern recognition, classification and the optimization technique problems as due to parallel, iterative.

IV. EVALUATION PARAMETER

To test our result this work use following measures the accuracy of the, that is to say Precision, Recall and F-score. These parameters are depend on the TP, TN, FP and FN [1, 2, 13].

$$\Pr ecision = \frac{True\_Positive}{True\_Positive + False\_Positive}$$

$$\mathrm{Re}call = \frac{True\_Positive}{True\_Positive + False\_Negative}$$

$$F\_Score = \frac{2 * \Pr ecision * \mathrm{Re}call}{\Pr ecision + \mathrm{Re}call}$$

In above true positive value is obtain by the system when the session is intrusion and actual class also says that session is intrusion. While in case of false positive value it is obtain by the system when the input session is not intrusion and actual class also says that session is intrusion.

V. CONCLUSION

Detection of intrusion in a network is an important issue as number of researchers have proposed different model for its detection. This paper present types of attack present in the neural network with their classes and different approaches of attacks perform by intruder. Various researcher work was also summarized with there approaches for detection of intrusion. In each approach training of machine is common, so detection accuracy of intrusion depends on learning model. Further paper has shown feature reduction methods for reducing the training and testing dataset size. Finally evaluation parameters which were common in most of research paper for comparison of intrusion detection techniques was shown. In future scholars should propose a model which can reduce the training feature of intrusion dataset.

REFERENCES

1. R. Karthik, Dr.S.Veni, Dr.B.L.Shivakumar "Improved Extreme Learning Machine (IELM) Classifier For Intrusion Detection System" International Journal of Engineering Trends and Technology (IJETT) – Volume-41 Number-2 - November 2016
2. Premansu sekhara rath, 2manisha mohanty, 3silva acharya, 4monica aich "optimization of ids algorithms using data mining technique" International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-4, Issue-3, Mar.-2016
3. Mohammadreza Ektefa, Sara Memar, Fatimah Sidi, Lilly Suriani Affendey "Intrusion Detection Using Data Mining Techniques", 978-1-4244-5651-2/10/$26.00 ©2010 IEEE
4. YU-XIN MENG," The Practice on Using Machine Learning For Network Anomaly Intrusion Detection" Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, 978-1-4577-0308-9/11/$26.00 ©2011 IEEE
5. S. Latha and S. J. Prakash, "A survey on network attacks and Intrusion detection systems," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, 2017, pp. 1-7.
6. K. Park, Y. Song and Y. Cheong, "Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm," 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService), Bamberg, 2018, pp. 282-286.
7. A. Nisioti, A. Mylonas, P. D. Yoo and V. Katos, "From Intrusion Detection to Attacker Attribution: A Comprehensive Survey of Unsupervised Methods," in IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 3369-3388, Fourthquarter 2018, doi: 10.1109/COMST.2018.2854724.
8. Koushal Kumar, Jaspreet Singh Batth "Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 150 – No.12, September 2016.
9. Yogita B. Bhavasar, Kalyani C. Waghmare "Intrusion Detection System Using Data Mining Technique: Support Vector Machine" 2013 International Journal of Emerging Technology and Advance Engineering volume 3, Issue 3, March 2013.
10. A.R. Jakhale, G.A. Patil, "Anomaly Detection System by Mining Frequent Pattern using Data Mining Algorithm from Network Flow", International Journal of Engineering Research and Technology, Vol. 3, No.1, January 2014, ISSN. 2278-0181.
11. Barolli, Leonard; Elmazi, Donald; Ishitaki, Oda, Tetsuya; Taro; Yi Liu, Uchida, Kazunori. (24-27 March 2015). Application of Neural Networks for Intrusion Detection in Tor Networks. Advanced Information Networking and Applications Workshops (WAINA), 2015 IEEE 29th International Conference on, p67-72.
12. KAIYUAN JIANG , WENYA WANG , AILI WANG , AND HAIBIN WU. "Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network". IEEE Access Volume 8, February 24, 2020.
13. Chuanlong Yin, Yuefei Zhu, Jinlong Fei, And Xinzheng He. "A Deep Learning Approach For Intrusion Detection Using Recurrent Neural Networks" current version November 7, 2017. *Digital Object Identifier 10.1109/ACCESS.2017.2762418*

14. Govinda.K1, Kevin Thomas. "Survey on Feature Selection and Dimensionality Reduction Techniques". International Research Journal of Engineering and Technology Volume: 03 Issue: 07, 2016.
15. Abdi, H. Salkind, N. (ed.) Encyclopedia of measurements and statistics Singular value decomposition (SVD) and Generalized Singular Value Decomposition (GSVD) Sage Publications, 2007, 907-912.
16. S. Khalid, T. Khalil and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," 2014 Science and Information Conference, London, 2014, pp. 372-378, doi: 10.1109/SAI.2014.6918213.