International Journal of Electrical and
Electronics Engineering (IJEEE)
ISSN(P): 2278–9944; ISSN(E): 2278–9952
Vol. 8, Issue 6, Oct–Nov 2019, 1–10

International Academy of Science,
Engineering and Technology
Connecting Researchers; Nurturing Innovations

INTRUSION DETECTION SYSTEM TO MITIGATE CYBER ATTACKS USING MACHINE LEARNING

Bhavika G. S[1] & Jagadeesh Sai D[2]

[1]Research Scholar, Department of Information Science, Ramaiah Institute of Technology, Bengaluru, Karnataka, India

[2]Assistant Professor, Department of Information Science, Ramaiah Institute of Technology, Bengaluru, Karnataka, India

ABSTRACT

Cyber attacks which are malicious and deliberate are one of the major ultimatums that hit the business, organizations and institutions every day. Despite of having various cyber defense systems, Web Application Firewalls, intrusions are the common threats that exist till today. Intrusion detection systems are the new generation security technologies and they detect the known attacks and also the unknown after the system is affected by an attack. The need for predicting the detection accuracy of an attack is one of the major concerns. Here we propose an intrusion detection model that predicts the accuracy of attack detection and performance is evaluated on NSL-KDD, an effective benchmark dataset using machine learning technique.

KEYWORDS: *Intrusion Prediction, Machine Learning, Cyber Attacks, Intrusion Detection Systems, Random Forest*

Article History

Received: *03 Oct 2019* | ***Revised:*** *15 Oct 2019* | ***Accepted:*** *23 Oct 2019*

INTRODUCTION

Cyber attacks strike the businesses every day. They have the potential to evade the network and hack the useful information of an origination or an individual by getting unauthorized access to their systems. This is nothing but the intrusion and an attempt made to exploit the confidentiality of an organization or an individual. Web Application Firewalls are one of the technologies used to mitigate cyber attacks by filtering the http requests. WAF is a hardware or software used in between the networks to mitigate the attacks by using some policies. These policies will safeguard the system by vulnerabilities by filtering the malicious content. It's been clear that firewalls alone are not enough to secure the network as it does not counter the attacks occurred outside the system or network.

Intrusion Detection System (IDS) serves better for this existing challenge of WAF and is considered as a new generation of security technology. IDS mainly focus on detecting the attacks rather than prevention. IDS is extensively used as it overcomes the major drawback of WAF. This paves the way to new technology, IDS can be classified into mainly two streams; anomaly based intrusion detection system and misuse or signature based intrusion detection system. Misuse based IDS is used extensively where there is less risk of cyber attacks as it detects only the known or predefined or programmed attacks. Anomaly based IDS system is kept under the observation to figure out the behavioral profile of the system. Whenever a deviation occurs beyond the normal behavior it is flagged as attack.

Machine learning is capable of extracting the patterns of attacks or intrusions, to build a normal behavior of a system, to classify the attacks, to gain high accuracy in attack detection and mainly in assisting the security experts to analyze the pattern of attacks. In this paper, we present the application of machine learning to intrusion detection. NSL-KDD dataset has been used and performance is evaluated by different metrics like accuracy, specificity, and sensitivity and computation time.

Types of Cyber Attacks

Many researchers and experts have termed or defined cyber attack as a malicious code that can evade the computer network or system that has the capability to exploit the user information and used by hackers for offensive work. Some of the common types of cyber attacks are as follows:

Malware: It is the software which is specially designed to disrupt the network and it describes ransome, worms, spyware and viruses. Malware contraventions a network through vulnerability, when a user opens a malicious mail or clicking an unsafe link.

- Some key component access will be lost or blocked.
- Harmful software or a malware will be installed in the system.
- Spyware covertly obtains data into the system from the hardware.
- Makes the system inoperable.

Phishing: It is sending a fraudulent messages or suspicious links through email as if it is coming from a reputed source. This type of practice is said to be phishing attack and one of the common vulnerabilities. The main motive here is to steal the sensitive information like debit or credit passwords and others.

Man-in-the-Middle Attack: This is also known as Eaves-dropping, occurs when the third person or attacker interfere a two-party transaction or communication. The sender and receiver will never come to know about the man in the middle and they exchange information. This information will be recorded by the attacker and used for offensive purposes.

Daniel-of-Service-Attack (DOS): we usually come across situations like mailbox or a server flooded with lump sum of requests. This kind of situations is DOS attacks where the attacker launches the enormous requests to exhaust the resources of the network bandwidth.

Sql Injection: when a server gets injected by a malicious code that uses SQL, it forces the server to reveal or open up the information to the attacker. This kind of vulnerability is termed as SQL injections.

Zero-Day-Exploits: it requires constant consciousness. Here the attack is performed before the solution to vulnerability is implemented by the server or the network.

U2r Attack: User to root attack is one of the widely known vulnerabilities and tries to attack the entire network when accessing it legally.

R2l Attack: Remote to Local attack is also similar to U2R attack, it creates vulnerability by hacking the entire network by gaining unauthorized access to network.

Probe: it collects information about the network activity by inserting a malicious program or a device at a key point of network. Through a weak point in the computer system probe attack is evaded and gains access to network.

METHODOLOGY

Main objective of the proposed model is to improve the attack detection accuracy using machine learning. Here we incorporate Support Vector Machine algorithm and Random Forest Model; finally we compare both the results in terms of accuracy using NSL-KDD dataset.

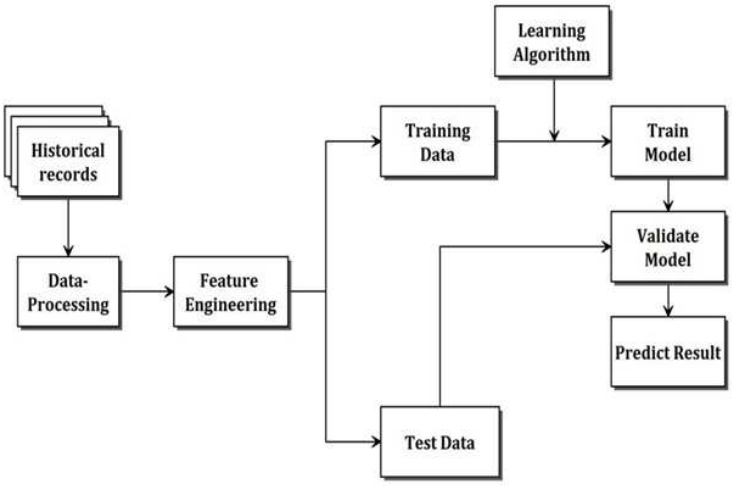

Figure 1: Intrusion Detection Framework Architecture.

Dataset Description

- The dataset consists of no redundant values so that classifier algorithm works well to distinguish the attacks and does not give biased result.
- Test set does not consist of replicated data and yields better reduction rates.
- The dataset consist of 10,000 rows and 43 columns. Test data is taken as 30% and remaining is trained.

The model is based on Machine Learning concept, where the data is preprocessed and split into training and testing dataset to get better accuracy. Support Vector Machine and Random Forest Algorithm Models are implemented and compared for highest detection accuracy. Random Forest yields better precision and accuracy over SVM.

Support Vector Machine

SVM is a binary classifier algorithm based on statistical learning technique for both regression and classification tasks. It has a wide variety of applications and recent advancements are its use in information security. The two important advantages of SVM are its ability of overcoming the curse of dimensionality and generalisation nature. One of the main reasons to use SVM in intrusion detection is its ability to detect the attacks or intrusions. Despite of its advantages SVM has some drawbacks when it is applied in IDS. Being a supervised machine learning algorithm its more time consuming, requires labelled data for efficient learning.

Random Forest Model

Random decision forest or random forest algorithm is also known as ensemble learning model. It is used for both regression and classification tasks as it functions by building multiple decision trees at the training period of data and gives the output as mean prediction class.

Random forest is considered to be the best approach in intrusion detection as it is easy to use, good prediction results are often obtained by hyper-parameters and it yields 99% accuracy in attack detection.

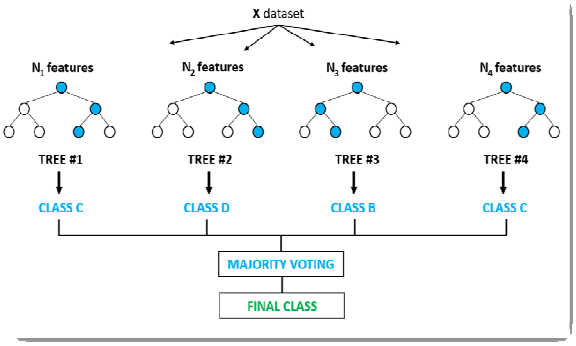

Figure 2: Random Forest Model.

LITERATURE REVIEW

Author states that [1] with the increase in severity and number of attacks; information flourishing out day by day on internet, World Wide Web has become a place for attackers. And there comes a strong need for securing data and preventing intrusions. In this paper, they have analysed the http requests to know the web attacks pattern and design a CNN to detect the attacks and evaluate the performance.

In this paper [2] author discusses the importance of Machine learning and Deep Learning Models in intrusion detection. Earliest attack detection techniques like firewalls are not enough to safeguard the target servers and systems. They propose deep learning models produces better accuracy in attack detection using CSIC 2010 dataset for evaluation.

According to [3] Optimising the performance of IDS due to dynamic properties of data and intrusion behaviours, need for securing and improvising the existing IDS is a major requirement in information security today. Machine learning approach helps in improvising the performance of IDS in efficient manner.

As stated in [4] Web Applications and Web servers are the common venue for the attackers and today it resides as a global threat in information security. HTTP requests are the main form of attacking strategy attackers always use. Flourishing the user's mailbox with hundreds of junk mails and malicious links, sending malicious http requests are the various ways of attacking. By collecting the normal behaviour of http logs and constructing a normal profile of the system using feature extraction.

In [5] a recurrent neural network sequence-to-sequence system call has been proposed to overcome the drawback of IDS that are not efficient to predict the behaviour of the system. Using the trained model and system calls sequence, we can predict the intrusion behaviour. By improvising the system calls sequence anomaly detection algorithm's efficiency can be increased.

In this research [6] authors states that many research conducted on intrusion detection have made use of various benchmark dataset like DARPA, KDD Cup 99, etc. Drawbacks of these datasets has been overcome by NSL-KDD dataset and is widely been used in intrusion detection. This paper mainly deals with four attributes of dataset, traffic, content, host and basic. The empirical analysis of these attributes results in sustainable dataset to achieve less false alarm rates and high accuracy.

Here [7] the authors mainly focuses on four major types of attacks which the IDS deals with, DoS, Probe, r2l, U2r, normal. Among these they made a consistent study on U2R attacks User-to-Root Attacks which leads many malicious functions like dictionary attack, social engineering threats and sniffing passwords. Many learning models like J48, Random forest, Naïve Bayes, Multiple preceptron are used to achieve accuracy.

In this paper [8] advancements in online services and vulnerabilities attached to these services are discussed. In order to discover the newly caused threats, web logs from the web servers are collected to build a normal behavior of the system so that any deviations from the built behavior can be flagged as an attack or threat.

Author has made a comparative study on Classification algorithms for intrusion detection [9]. Top ten classification models have been compared namely, J48, Logistic Regression, PART, SGD, BayesNet, IBK, Random Forest and REPTree. Experimenting with these top classification models in WEKA environment, the author has concluded that decision tree algorithm comes out as the best classifier. RF model comes out as high accuracy predictor and IBK takes less computational time. By implementing these IDS may work better and gain high performance.

Using Novel approach of Honey Tokens [10] author has proposed an IDS model to secure critical infrastructure networks. Encrypted pointers and honey tokens in the frame helps in trapping the attacker, by dividing the network into four different pools or divisions. All these pools incorporate different encryption algorithms like AES-128, AES-192, and AES-256, etc to reduce the false alarm rates in IDS. Experimental results are shown in realistic conditions.

Offering a secured environment or path for web Applications are not easy task. Not only a secured environment serves web app to be safe but the other main parameters like up-to-date shared hosts, system configurations also matters. However Web Application Firewalls tries to fill this gap, it fails to do so. In this paper [11] an anomaly based intrusion detection model is proposed to safe web applications.

EXPERIMENTS AND RESULTS

Experiments are performed using python in Jupiter Notebook using 30% of the NSL-KDD dataset as testing data. First data pre-processing and removal of null values in data are performed. Data analysis is done by plotting the xAttack variable with different attributes like dst_host_rerror_rate, network service on the destination, dst_host_same_src_port_rate, and srv_serror_rate and protocol type.

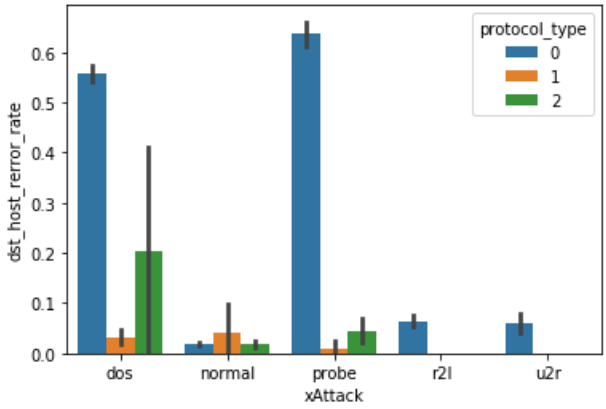

Figure 3: Barplot of Xattack Vs Dst_Host_Rerror_Rate.

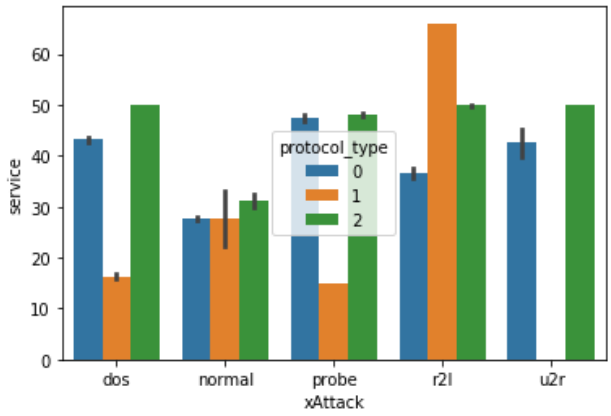

Figure4: Bar Plot of Xattack Vs Service.

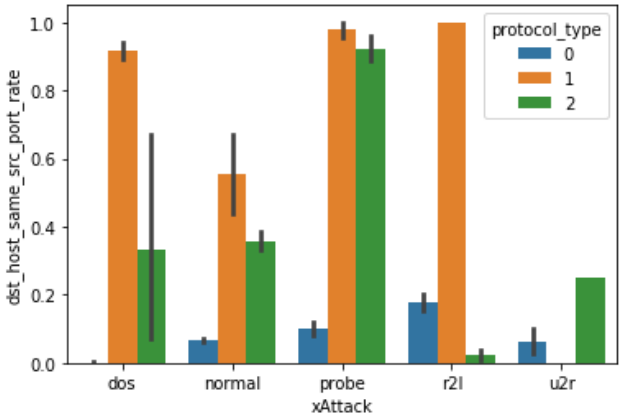

Figure 5: Bar Plot of Xattack Vs Dst_Host_ Same_Src_Port_Rate.

Accuracy of Support Vector Machine is evaluated by one of the performance metric Confusion matrix.

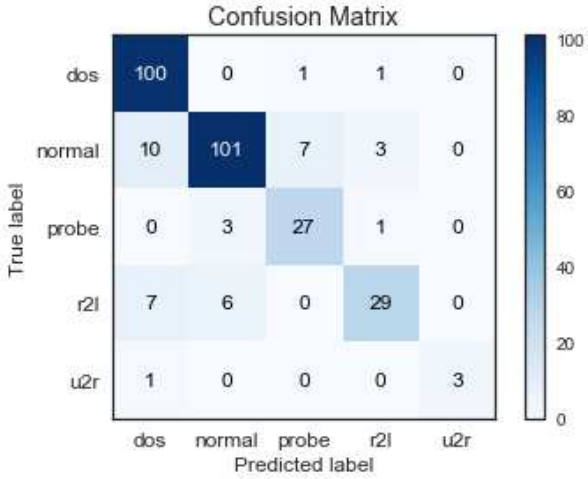

Figure 6: Confusion Matrix For SVM.

Similarly Random forest mode is also evaluated by the performance metric Confusion matrix.

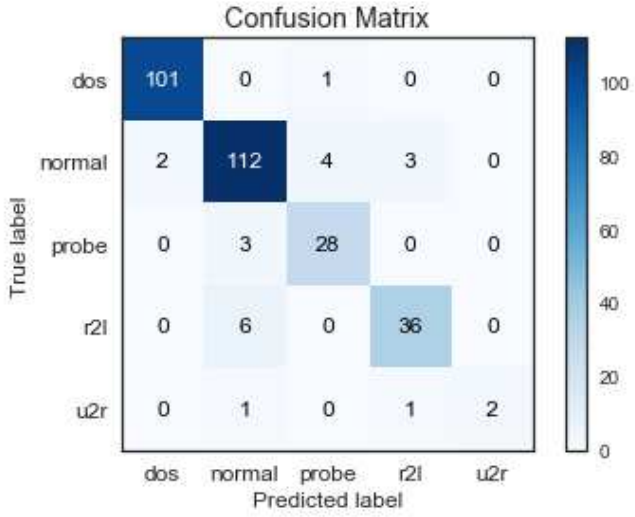

Figure 7: Confusion Matrix for RF Model.

Then we apply support Vector Machine to calculate the accuracy and precision of the model which yields 87% accuracy for the test data.

Finally, we use the Random Forest model to predict the accuracy of the model and it yields 93% for the test data. Comparison of both models is plotted using bar plots.

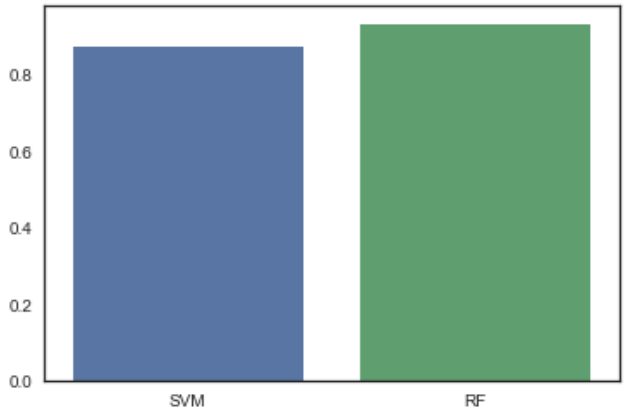

Figure 8: Comparison of SVM and RF Model.

CONCLUSIONS

Over the past few years, increasing in the number of attacks and its severity, need for improving the existing IDS is one of the major tasks to be carried out in Information Security. So we presented the design and implementation of efficient IDS by using Random Forest Model to overcome the existing drawbacks of IDS and we try to improve the attack detection accuracy of the system. A Django interface is created which shows the accuracy prediction of the system for various attacks for better visualisation and understanding of the system.

ACKNOWLEDGMENT

I mainly thanks to my guide Mr. Jagadeesh Sai D who has provided the extensive guidance and expertise throughout the study. Thanks to all who helped me in completing this work.

REFERENCES

1. *Ming Zhang(&), Boyi Xu, Shuai Bai, Shuaibing Lu, and Zhechao Lin, "A Deep Learning Method to Detect Web Attacks Using a Specially Designed CNN", Springer International Publishing AG 2017 D. Liu et al. (Eds.): ICONIP 2017, Part V, LNCS 10638, pp. 828–836, 2017. https://doi.org/10.1007/978-3-319-70139-4_84*
2. *Ali Moradi Vartouni, Saeed Sedighian, Mohammed Teshnehlab, "An Anomaly Detection Method to Detect Web Attacks Using Stacked Auto-Encoder", 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*
3. *Jayshree Jha, Leena Ragha, Ph.D, "Intrusion Detection System using Support Vector Machine", International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA International Conference & workshop on Advanced Computing 2013 (ICWAC 2013) – www.ijais.org.*
4. *Mikhail Zolutukhin, timo Hamalainen, Tero Kokkonen, Jarmo Siltanen, "Analysis of HTTP requests for anomaly detection of web attacks", 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing.*
5. *Shaohua lv, jian wang, yinqi yang, and jiqiang liu, "Intrusion Prediction With System-Call Sequence-to-Sequence Model", Received September 23, 2018, accepted November 4, 2018, date of publication November 19, 2018, date of current version December 18, 2018.*
6. *Preeti Aggarwala,*, Sudhir Kumar Sharmab, "Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection", 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015) Elsevier Publicatiions.*
7. *S. Revathi 1 Dr. A. Malathi2, "Detecting User-To-Root (U2R) Attacks Based on Various Machine Learning Techniques", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 4, April 2014.*
8. *Yang Goa, Yang Ma, "Anomaly Detection of Malicious Users' Behaviors for Web Applications Based on Web Logs" 2017 17th IEEE International Conference on Communication Technology.*
9. *Himadri Chauhan, Vipin Kumar, Sumit Pundir, Emmanuel S. Pilli, "A Comparative Study of Classification techniques for Intrusion Detection ", 2013 IEEE, International Symposium on Computational and Business Intelligence.*
10. *Muhammad Kamran Asif, Yahya Subhi Al-harthi "Intrusion Dtection System using Honey Token based Encrypted Pointers to Mitigate Cyber Threats for Critical Infrastructure Networks", 2014 IEEE International Conference on Systems, Man and Cybernetics.*
11. *Muhammad Kamran Asif, Yahya Subhi Al-harthi "Anomaly Detection for Web server log rreduction: a simple yet efficient crwaling based approach", The 2nd IEEE Workshop on Security and Privacy in the cloud (SPC 2016).*

AUTHOR PROFILE

Bhavika G.S is M.Tech degree holder from Ramaiah Institute of technology in Information Science, Software Engineering. Presently working as Software Engineer in KPIT Technologies Bengaluru.

Jagadeesh Sai D. is a M.Tech degree holder and is serving as an assistant professor in IS&E department of MSRIT. He is interested in subjects related to IoT ,Programming lanugages ,Scritping,Cryptographyand Distributed Systemsoperating system.