International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET122939

Network attack Prediction using Supervised ML Algorithm

R. Geetha[1], Mr. Shubham Burman[2], Vikram Posala[2], Pavan Kumar S[2], Sagar K K[2]

[1]Assistant Professor, CiTech, Bangalore, Karnataka, India

[2]CiTech, Bangalore, Karnataka, India

Article Info
Volume 9, Issue 3
Page Number : 190-195

Publication Issue :
May-June-2022

Article History
Accepted : 10 May 2022
Published: 20 May 2022

ABSTRACT

Generally, to create data for the Intrusion Detection System (IDS), it is necessary to set the real working environment to explore all the possibilities of attacks, which is expensive. Software to detect network intrusions protects a computer network from unauthorized users, including perhaps insiders. The intrusion detector learning task is to build a predictive model (i.e. a classifier) capable of distinguishing between "bad" connections, called intrusions or attacks, and "good" normal connections. To prevent this problem in network sectors have to predict whether the connection is attacked or not from Kaggle dataset using machine learning techniques. The aim is to investigate machine learning based techniques for better packet connection transfers forecasting by prediction results in best accuracy. To propose a machine learning-based method to accurately predict the DOS, R2L, UU2R, Probe and overall attacks by prediction results in the form of best accuracy from comparing supervise classification machine learning algorithms. Additionally, to compare and discuss the performance of various machine learning algorithms from the given dataset with evaluation classification report, identify the confusion matrix and to categorizing data from priority and the result shows that the effectiveness of the proposed machine learning algorithm technique can be compared with best accuracy with precision.

Keywords : DOS, R2L, UU2R

I. INTRODUCTION

Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine learning algorithm using python. Process of training and prediction involves use of specialized algorithms. It feed the training data to an algorithm, and the algorithm uses this training data to give predictions on

a new test data. Machine learning can be roughly separated in to three categories[1]. There are supervised learning, unsupervised learning and reinforcement learning. Supervised learning program is both given the input data and the corresponding labeling to learn data has to be labeled by a human being beforehand. Unsupervised learning is no labels. It provided to the learning algorithm. This algorithm has to figure out the clustering of the input data. Finally, Reinforcement learning dynamically interacts with its environment and it receives positive or negative feedback to improve its performance.

Data scientists use many different kinds of machine learning algorithms to discover patterns in python that lead to actionable insights. At a high level, these different algorithms can be classified into two groups based on the way they "learn" about data to make predictions: supervised and unsupervised learning. Classification is the process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function from input variables(X) to discrete output variables(y). In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. This data set may simply be bi-class (like identifying whether the person is male or female or that the mail is spam or non-spam) or it may be multi-class too. Some examples of classification problems are: speech recognition, handwriting recognition, bio metric identification, document classification etc [2].

Supervised Machine Learning **is the** majority of practical machine learning uses supervised learning. Supervised learning is where have input variables (X) and an output variable (y) and use an algorithm to learn the mapping function from the input to the output **is y = f(X).** The goal is to approximate the mapping function so well that when you have new input data (X) that you can predict the output variables (y) for that data. Techniques of Supervised Machine Learning algorithms include **logistic regression**, **multi-class classification**, **Decision Trees** and **support vector machines etc**.

Supervised learning requires that the data used to train the algorithm is already labeled with correct answers. Supervised learning problems can be further grouped into **Classification** problems. This problem has as goal the construction of a succinct model that can predict the value of the dependent attribute from the attribute variables. The difference between the two tasks is the fact that the dependent attribute is numerical for categorical for classification [3]. A classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model will try to predict the value of one or more outcomes. A classification problem is when the output variable is a category, such as "red" or "blue"processing. The attendance report also can be sent through email to the required individual, if required [1]. Attendance marking in a classroom during a lecture is not only a onerous task but also a time consuming one at that. Due to an unusually high number of students present during the lecture there will always be a probability of proxy attendance. Attendance marking with conventional methods has been an area of challenge. The growing need of efficient and automatic techniques of marking attendance is a growing challenge in the area of face recognition. In recent years, the problem of automatic attendance marking has been widely addressed throughthe use of standard biometrics like fingerprint and Radio frequency Identification tags etc., However, these techniques lack the element of reliability [4]. In this proposed project an automated attendance marking and management system is proposed by making use of face detection and recognition algorithms. Instead of using the conventional methods, this proposed system aims to develop an automated system that records the student's attendance by using facial recognition technology.

II. LITERARTURE REVIEW

1. THE FEASIBILITY OF LAUNCHING AND DETECTING JAMMING ATTACKS IN WIRELESS NETWORKS

AUTHORS: W. Xu, W. Trappe, Y. Zhang, and T. Wood

Wireless networks are built upon a shared medium that makes it easy for adversaries to launch jamming-style attacks. These attacks can be easily accomplished by an adversary emitting radio frequency signals that do not follow an underlying MAC protocol. Jamming attacks can severely interfere with the normal operation of wireless networks and, consequently, mechanisms are needed that can cope with jamming attacks. In this paper, we examine radio interference attacks from both sides of the issue: first, we study the problem of conducting radio interference attacks on wireless networks, and second we examine the critical issue of diagnosing the presence of jamming attacks. Specifically, we propose four different jamming attack models that can be used by an adversary to disable the operation of a wireless network, and evaluate their effectiveness in terms of how each method affects the ability of a wireless node to send and receive packets. We then discuss different measurements that serve as the basis for detecting a jamming attack, and explore scenarios where each measurement by itself is not enough to reliably classify the presence of a jamming attack. In particular, we observe that signal strength and carrier sensing time are unable to conclusively detect the presence of a jammer.

2. AN INTEGRATED APPROACH FOR JAMMER DETECTION USING SOFTWARE DEFINED RADIO

AUTHORS: R. Bhojani and R. Joshi

Due to shared nature of wireless communication any malicious user can easily monitored communication between two devices and emits false message to block communication. Nowadays increased use of software defined radio (SDR) technology makes any types of jammer device using same hardware with little modification in software. A jammer transmits radio signal to block legitimate communication either overlapping signal with more power or reducing signal to noise ratio. In this paper we have survey different jammer detection methods for efficient detection of jammers presence in system. Existing jammer detection methods like packet delivery ratio (PDR) packet send ratio (PSR), bad packet ratio (BPR) and signal to noise ratio (SNR) can effectively detects jammer, here we have proposed novel method for jammer detection using communication parameter used in SDR like synchronization indicator, iteration and adaptive signal to jammer plus noise ratio (ASNJR).

3. AN SVM-BASED METHODFOR CLASSIFICATION OF EXTERNAL INTERFERENCE IN INDUSTRIAL WIRELESS SENSOR AND ACTUATOR NETWORKS

AUTHORS : S. Grimaldi, A. Mahmood, and M. Gidlund

In recent years, the adoption of industrial wireless sensor and actuator networks (IWSANs) has greatly increased. However, the time-critical performance of IWSANs is considerably affected by external sources of interference. In particular, when an IEEE 802.11 network is coexisting in the same environment, a significant drop in communication reliability is observed. This, in turn, represents one of the main challenges for a wide-scale adoption of IWSAN. Interference classification through spectrum sensing is a possible step towards interference mitigation, but the long sampling window required by many of the approaches in the literature undermines their run-time applicability in time-slotted channel hopping (TSCH)-based IWSAN.

4. A TRAINING ALGORITHM FOR OPTIMAL MARGIN CLASSIFIERS

AUTHORS : B. E. Boser, I. M. Guyon, and V. N. Vapnik

ABSTRACT

A training algorithm that maximizes the margin between the training patterns and the decision

boundary is presented. The technique is applicable to a wide variety of the classification functions, including Perceptrons, polynomials, and Radial Basis Functions. The effective number of parameters is adjusted automatically to match the complexity of the problem. The solution is expressed as a linear combination of supporting patterns. These are the subset of training patterns that are closest to the decision boundary. Bounds on the generalization performance based on the leave-one-out method and the VC-dimension are given. Experimental results on optical character recognition problems demonstrate the good generalization obtained when compared with other learning algorithms.

III. SYSTEM ANALYSIS

Analysis is the process of breaking a complex topic or substance into smaller parts to gain a better understanding of it. Gathering requirements is the main attraction of the Analysis Phase. The process of gathering requirements is usually more than simply asking the users what they need and writing their answers down. Depending on the complexity of the application, the process for gathering requirements has a clearly defined process of its own [5].

Existing System

The system focuses on the conception of a monitoring network that can able to detect and classify jamming and protocol-based attacks. To achieve this goal, the system proposed to outsource the attack detection function to protect the network and used an antenna to monitor the spectrum over the time. The Wi-Fi network and the attacks were carried out in an anechoic chamber to avoid disturbing other Wi-Fi communication networks in the vicinity. The spectra highlights that the frequencies of interest belong to the communication channel between 2.402 and 2.422 GHz. Focusing the analysis on this 20-MHz frequency band permits to construct a classification model to overcome the problems induced by the utilization of the adjacent channels that can be or not occupied by other Wi-Fi communications. On these frequencies, the proposed estimation model shows good results in the prediction of attacks. In addition, the correction using the K spectra nearest in time permits to correct most of the miss classification[6].

IV. PROPOSED SYSTEM

Exploratory Data Analysis

This analysis is not meant to be providing a final conclusion on the reasons leading to network sector as it doesn't involve using any inferential statistics techniques/machine learning algorithms. Machine learning supervised classification algorithms will be used to give the network connection dataset and extract patterns, which would help in predicting the likely patient affected or not, thereby helping the attack of avoids for making better decisions in the future. Multiple datasets from different sources would be combined to form a generalized dataset, and then different machine learning algorithms would be applied to extract patterns and to obtain results with maximum accuracy [6][9].

Data Wrangling

In this section of the report will load in the data, check for cleanliness, and then trim and clean given dataset for analysis. Make sure that the document steps carefully and justify for cleaning decisions.

- **Data collection**

The data set collected for predicting the network attacks is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using Random Forest, logistic, Decision tree algorithms, K-Nearest Neighbor (KNN) and Support vector classifier (SVC) are applied on the Training set and based on the test result accuracy, Test set prediction is done.

Preprocessing

The data which was collected might contain missing values that may lead to inconsistency[10]. To gain better results data need to be preprocessed so as to improve the efficiency of the algorithm. The outliers have to be removed and also variable conversion need to be done. The correlation among attributes can be identified using plot diagram in data visualization process. Data preprocessing is the most time consuming phase of a data mining process. Data cleaning of connections, data removed several attributes that has no significance about the behavior of a packet transfers. Data integration, data reduction and data transformation are also to be applicable for network connections dataset. For easy analysis, the data is reduced to some minimum amount of records. Initially the Attributes which are critical to make a loan credibility prediction is identified with information gain as the attribute-evaluator and Ranker as the search-method [7][8].

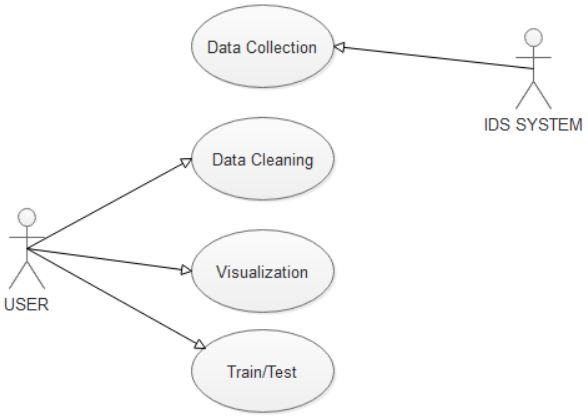
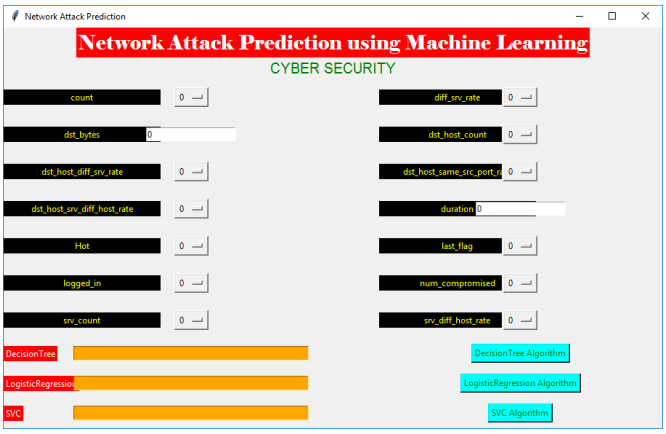
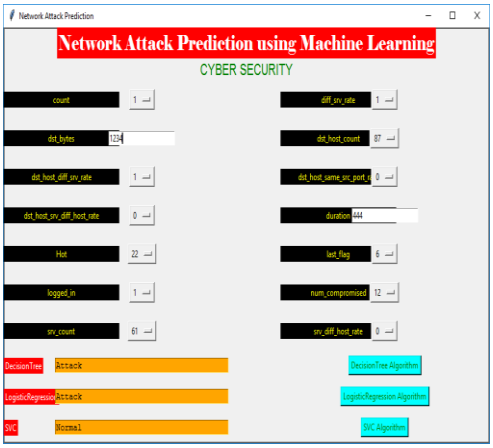


Fig: Use case Diagram

V. RESULTS





VI. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score of Decision tree algorithm which brings some insights about diagnosing network attack for new connection. A prediction model is presented with aid of AI to improve human accuracy and provide early detection scope. This model inferred that, area analysis and use of ML technique is useful in developing prediction models that can helps to network sectors reduce the long process of diagnosis and eradicate any human error. The outcome of this project is finding the best accuracy of network attacks using ML algorithm.

VII. REFERENCES

[1]. Bindra, Naveen & Sood, Manu. (2019), Detecting DDoS Attacks Using Machine Learning Techniques and Contemporary Intrusion Detection Dataset Automatic Control and Computer Sciences. 53. 419-428. 10.3103/S0146411619050043.

[2]. M. Almseidin, M. Alzubi, S. Kovacs and M. Alkasassbeh, (2017), "Evaluation of machine learning algorithms for intrusion detection system," IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, 2017, pp. 000277- 000282.

[3]. Mellor, A., Haywood, A., Stone, C., and Jones, S., (2013) The performance of random forests in an operational setting for large area sclerophyll forest classification, Remote Sens., vol. 5, no. 6, pp. 2838–2856.

[4]. Arul, Amudha & Subburathinam, Karthik & Sivakumari, S. (2013). Classification Techniques for Intrusion Detection - An Overview. International Journal of Computer Applications. 76. 33-40. 10.5120/13334-0928.

[5]. Kanagalakshmi. R, V. Naveenantony Raj, (2014) Network Intrusion Detection Using Hidden Naïve Bayes Multiclass Classifier Model, International Journal of Science, Technology & Management ,Volume No.03, Issue No. 12.

[6]. M. Alkasassbeh, G. Al-Naymat et.al, (2016) Detecting Distributed Denial of Service Attacks Using Data Mining Technique,' (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, pp. 436-445.

[7]. Jasreena Kaur Bains ,Kiran Kumar Kaki ,Kapil Sharma, (2013) Intrusion Detection System with Multilayer using Bayesian Networks, International Journal of Computer Applications (0975 – 8887) Volume 67– No.5.

[8]. Dewan Md. Farid, Nouria Harbi, Mohammad Zahidur Rahman, (2010) Combining Naive Bayes and Decision Tree for Adaptive Intrusion Detection, Proc. of Intl. Journal of Network Security & Its Applications (IJNSA), Volume 2, pp.12-25.

[9]. Domingos P. and Pazzani M., Beyond Independence: Conditions for the optimality of the simple Bayesian Classifier, in proceedings of the 13th Intnl. Conference on Machine Learning, 1996, pp.105-110.

[10]. V. Hema and C. Emilin Shyni, (2015) DoS Attack Detection Based on Naive Bayes Classifier, Middle-East Journal of Scientific Research 23 (Sensing, Signal Processing and Security): 398-405.

Cite this article as :

R. Geetha, Shubham Burman, Vikram Posala, Pavan Kumar S, Sagar K K, "Network attack Prediction using Supervised ML Algorithm", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 3, pp. 190-195, May-June 2022. Available at doi : https://doi.org/10.32628/IJSRSET122939
Journal URL : https://ijsrset.com/IJSRSET122939