IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

INTELLIGENT WIRELESS WAN ENCROACHMENT DISCERNMENT USING MACHINE LEARNING TECHNIQUES

Mr.S.CHITRAPANDI[1] , Mrs.S.P.AUDLINE BEENA

PG Scholar[1], Department of Computer Science, SRI Muthukumaran Institute of Technology, India

Assistant Professor[2], Department of Computer Science, SRI Muthukumaran Institute of Technology, India

ABSTRACT

Network attacks pose a significant threat to the security and integrity of computer networks. The ability to predict and prevent these attacks is crucial for maintaining a secure network environment. Supervised machine learning techniques have emerged as effective tools for network attack prediction due to their ability to analyse large amounts of network data and identify patterns indicative of malicious activity. We present a comprehensive analysis of supervised machine learning techniques for the prediction of network attacks. We collect and pre-process the data, extracting relevant features and transforming them into a suitable format for machine learning algorithms.

We evaluate the performance of these algorithms. We investigate the interpretability of the trained models to gain insights into the underlying patterns and characteristics of network attacks. This allows network administrators to understand the nature of attacks and develop appropriate defences strategies. Additionally, we discuss the challenges and limitations associated with the application of supervised machine learning techniques in the domain of network attack prediction, such as the need for real-time analysis and the emergence of sophisticated evasion techniques.

INTRODUCTION

The most devastating and complicated attack in a wireless sensor network is the Wormhole attack. In this attack, the attacker keeps track of the packets and makes a tunnel with other nodes of different communication networks, and thus the attacker passes the packets through this tunnel. And the outsider attack can be prevented by authentication and encryption techniques by launching a Sybil attack on the sensor network. In WSN the routing protocols in network has a unique identity. The figure demonstrates

Sybil attack where an attacker node_AD' is present with multiple identities.

EXISTING SYSTEM

Distributed denial-of-service (DDoS) attacks continue to grow at a rapid rate plaguing Internet Service Providers (ISPs) and individuals in a stealthy way. Thus, intrusion detection systems (IDSs) must evolve to cope with these increasingly sophisticated and challenging security threats. Traditional IDSs are prone to zero-day attacks since they are usually signature-based detection systems. However, the lack of up-to-date labelled training datasets makes these ML/DL based IDSs inefficient. The privacy nature of these datasets and widespread emergence of adversarial attacks make it difficult for major organizations to share their sensitive data. Federated Learning (FL) is gaining momentum from both academia and industry as a new sub-field of ML that aims to train a global statistical model across multiple distributed users, referred to as collaborators, without sharing their private data. Due to its privacy-preserving nature, FL has the potential to enable privacy-aware learning between a large numbers of collaborators. This paper presents a novel framework, called MiTFed that allows multiple software defined networks (SDN) domains (*i.e.*, collaborators) to collaboratively build a global intrusion detection model without sharing their sensitive datasets. It a promising framework to cope with the new emerging security threats in SDN.

DISADVANTAGES

- In this paper a new collative comparison measure that reasonably represents the gains and losses due to encroachment discernment is proposed.
- Low accuracy parameters when compared with other discernment results.
- Deploy model is not available to use effectively in all over the domains.

PROPOSED SYSTEM:

We proposed a system to develop the project using machine learning algorithm. Recently, Machine learning and Artificial intelligence has plays a big role in various industries for their improvement and development. So we tried to implement machine learning algorithm to make them more securable. The aim of this project is about provide the thread to intimate the security to stop the thread before it impact huge loss to organization or individuals. We collect the previous record of the attacks that had happened over these times. By collecting these records our machine learning algorithm tried to find out the pattern to those dataset. After finding those patterns the machine is able to predict the instance based on previous records. By doing that with various algorithms we can get high accuracy. We say our model good based on high accuracy values.

SYSTEM ARCHITECTURE:

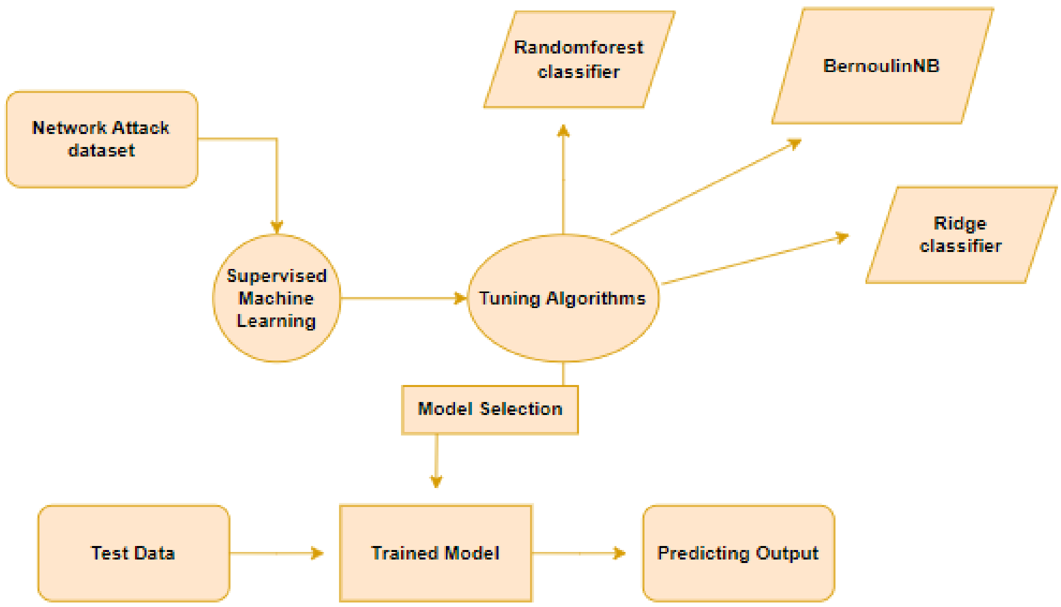


Fig . Architecture of this system

ADVANTAGES:

- It prevents the wireless attack from scammers by prior prediction.
- We build a framework based user friendly application using django.
- We use multiple machine learning algorithms for train data and Accuracy is improved.

VARIOUS ML ALGORITHMS COMPARED FOR ACCURACY:

It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple different machine learning algorithms in Python with scikit-learn.

In the example below 3 different algorithms are compared:

- BernoulinNB
- Randomforest Classifier
- Ridge Claassifier

Prediction result by accuracy:

Logistic regression algorithm also uses a linear equation with independent predictors to predict a value. The predicted value can be anywhere between negative infinity to positive infinity. It needs the output of the algorithm to be classified variable data. Higher accuracy predicting result is logistic regression model by comparing the best accuracy.

True Positive Rate(TPR) = TP / (TP + FN)

False Positive rate(FPR) = FP / (FP + TN)

Accuracy: The Proportion of the total number of predictions that is correct otherwise overall how often the model predicts correctly defaulters and non-defaulters.

Accuracy calculation:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same.

Precision: The proportion of positive predictions that are actually correct.

Precision = TP / (TP + FP)

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labeled as survived, how many actually survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

Recall: The proportion of positive observed values correctly predicted. (The proportion of actual defaulters that the model will correctly predict)

Recall = TP / (TP + FN)

Recall(Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it’s better to look at both Precision and Recall.

General Formula:

F- Measure = 2TP / (2TP + FP + FN)

F1-Score Formula:

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Result Analysis:

DATA VISUALIZATION AND DATA ANALYSIS

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df = pd.read_csv('WSN.csv')
df.head()
```

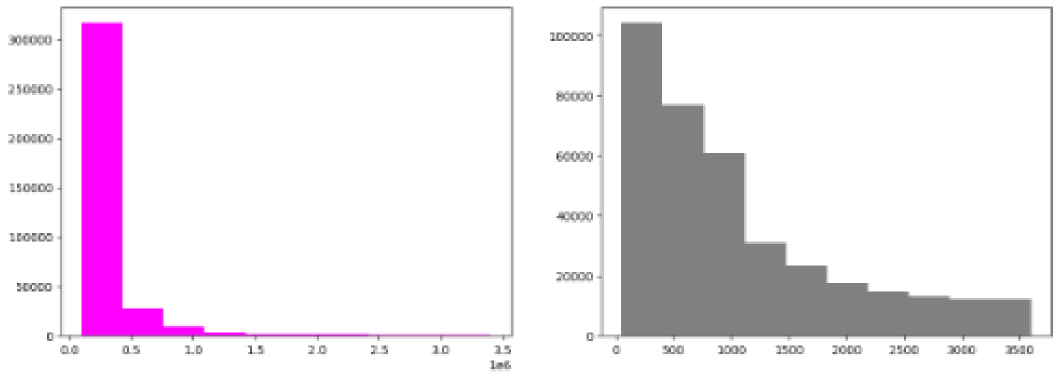
Out[2]:

	id	Time	Is_CH	who CH	Dist_To_CH	ADV_S	ADV_R	JOIN_S	JOIN_R	SCH_S	SCH_R	Rank	DATA_S	DATA_R	Data_Sent_To_BS	dist_CH_To_BS
0	101000	50	1	101000	0.00000	1	0	0	25	1	0	0	0	1200	48	130.08535
1	101001	50	0	101044	75.32345	0	4	1	0	0	1	2	38	0	0	0.00000
2	101002	50	0	101010	46.95453	0	4	1	0	0	1	19	41	0	0	0.00000
3	101003	50	0	101044	64.85231	0	4	1	0	0	1	16	38	0	0	0.00000
4	101004	50	0	101010	4.83341	0	4	1	0	0	1	25	41	0	0	0.00000

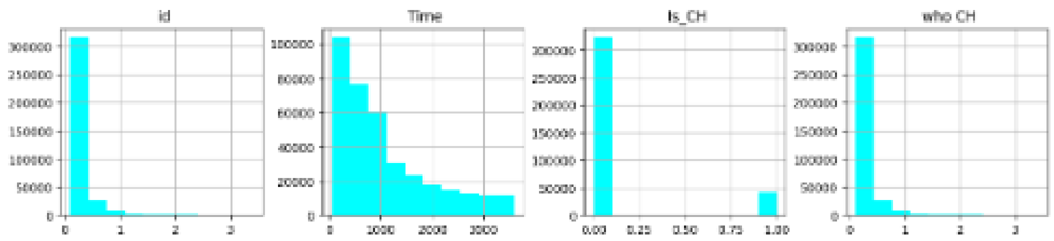
In [3]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

var = ['Attack_type']

for i in var:
    df[i] = le.fit_transform(df[i]).astype(int)
```

```
plt.figure(figsize=(15,5))

plt.subplot(1,2,1)
plt.hist(df[' who CH'],color='magenta')

plt.subplot(1,2,2)
plt.hist(df[' Time'],color='grey')
```

```
(array([103975.,  76962.,  60771.,  30923.,  23164.,  17747.,  14857.,
         13008.,  12237.,  12144.]),
 array([  50.,  405.,  760., 1115., 1470., 1825., 2180., 2535., 2890.,
        3245., 3600.]),
 <BarContainer object of 10 artists>)
```



```
df.hist(figsize=(15,55),layout=(15,4), color='cyan')
plt.show()
```



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 will be find out by comparing each algorithm with type of all WSN Attacks for future prediction results by finding best connections. This brings some of the following insights about diagnose the network attack of each new connection. To presented a prediction model with the aid of artificial intelligence to improve over human accuracy and provide with the scope of early detection. It can be inferred from this model that, area analysis and use of machine learning technique is useful in developing prediction models that can helps to network sectors reduce the long process of diagnosis and eradicate any human error.

REFERENCES:

[1] “Jinyin Chen , Jian Zhang, Zhi Chen, Min Du, and Qi Xuan, "Time-Aware Gradient Attack on Dynamic Network Link Prediction", *ieee transactions on knowledge and data engineering, vol. 35, no. 2, February 2023*.

[2] Dr. G. Umarani Srikanth and Priyadharsini.S, "Prediction Of Network Attacks using Machine Learning Techniques International Journal of Engineering Applied Sciences and Technology, *2021 Vol. 5, Issue 10, ISSN No. 2455-2143, Pages 112-118 Published Online February 2021 in IJEAST*.

[3] Zakaria Abou El Houda , Abdelhakim Senhaji Hafid , and Lyes Khoukhi, "MiTFed: A Privacy Preserving Collaborative Network Attack Mitigation Framework Based on Federated Learning Using SDN and Blockchain", *IEEE Transactions on network science and engineering, vol. 10, no. 4, July/August 2023*.

[4] Jakub Breier , Xiaolu Hou , Martı´n Ochoa , and Jesus Solano, "FooBaR: Fault Fooling Backdoor Attack on Neural Network Training", *IEEE Transactions on dependable and secure computing, vol. 20, no. 3, May/June 2023*.

[5] Qinghai Zhou , Liangyue Li , Nan Cao , Lei Ying , and Hanghang Tong, "Adversarial Attacks on Multi-Network Mining: Problem Definition and Fast Solutions", *IEEE Transactions on knowledge and data engineering, vol. 35, no. 1, January 2023*.

[6] K Pujitha; Gorla Nandini; K V Teja Sree; Banda Nandini; Dhodla Radhika, " Cyber Hacking Breaches Prediction and Detection Using Machine Learning", *2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies., vol. 12, no. 2, pp. 213-218, June 2023.*

[7] Aman Raj Pandey; Tushar Sharma; Subarna Basnet; Ankesh Kumar; Dr. Sonia Setia, " An Effective Phishing Site Prediction using Machine Learning", *2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)., 09 February 2023*.

[8] Pooja S Patil; S L Deshpande; Geeta S Hukkeri; R H Goudar; Poonam Siddarkar, "Prediction of DDoS Flooding Attack using Machine Learning Models", *2022 Third International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)., 16-17 December 2022*.