Original Article

Detection of Cyber Attack in Network Using different Machine Learning Approaches

B. Reddy Bharath[1], G. Yaswanth[2], Dr.C. Santhankrishnan[3]

[1]Department of Computer Science and Engineering, SRMIST, Chennai, India. E-mail: bb1481@srmist.edu.in

[2]Department of Computer Science and Engineering, SRMIST, Chennai, India. E-mail: gz2218@srmist.edu.in

[3]Department of Computer Science and Engineering, SRMIST, Chennai, India. E-mail: santhanc@srmist.edu.in

Abstract

In contrast to the past, advancements in computer and communication technologies have resulted in broad and rapid transformations. People, organizations, and governments all benefit from new inventions, yet some people, organizations, and governments are harmed. For example, security of important data, the security of stored information, and the accessibility of the data, among other things. Digital fear is one of the most critical challenges in today's world, based on these difficulties. Digital apprehension, which has caused a slew of concerns for individuals and organizations, has reached a point where it might jeopardize open and national security by many groups, including criminal organizations, Intrusion Detection Systems (IDS) were developed in this vein to keep a strategic distance from digital attacks. Currently, support vector machine computations were used to detect port sweep initiatives based on the new CICIDS 2017.

In the this paper we use the ensembled based hybrid classification which can enhance the better detection rate since here we use weak and strong classifiers.

Index Terms: SVM, Random Forest, Linear Regression, DDOS.

DOI: 10.47750/pnr.2022.13.S03.235

INTRODUCTION

Political and economic actors are increasingly using sophisticated cyber-warfare to disrupt, destroy, or suppress information content in computer networks. There is a requirement to assure network protocol resilience against incursions by powerful attackers who can even control a percentage of the network's parties. Both passive (eavesdropping, nonparticipation) and active (jamming, message dropping, corruption, and forging) assaults can be launched by the controlled parties. Intrusion detection is the system which continuously monitoring events in a computer system or network, analysing them for signals of potential problems, and, in many cases, preventing unwanted access. This is usually performed by automatically gathering data from a range of systems and network for potential security issues. Traditional intrusion detection and solutions, such as firewalls, access controlling mechanisms, and encryptions, have significant flaws when it comes to properly defending networks and systems against more complex assaults such as denial of service. Furthermore, most systems based on such methodologies have a high rate of false positive and false negative detection, as well as a lack of ability to react to changing harmful behaviour. Several Machine Learning (ML) approaches have, however, been applied to the challenge of intrusion detection in the last decade in the hopes of boosting detection rates and adaptability. These methods are frequently employed to maintain attack information bases current and thorough. Cyber-security and defence against a variety of cyber-attacks has recently become a hot topic. The fundamental reason for this is the phenomenal expansion of computer technology. a large number of relevant apps used by people or groups for personal or commercial purposes, particularly after the Internet of Things was accepted (IoT). The cyber-threats wreak havoc and generate significant financial losses on a huge scale networks. Hardware and software solutions that are already in place Firewalls, user authentication, and data encryption mechanisms are all examples of security measures. Not enough to address the anticipated demand problem, and Unfortunately, the computer network's multiple computers were unable to be protected. Cyber-threats. These traditional security arrangements aren't working. Sufficient as a protection as a result of the more rapid and rigorous evolution of intrusion detection systems Only the access from the firewall is controlled. The term "network to network" refers to the inability of two networks to communicate with each other. Networks. However, it does not send out any alerts in the event of an emergency. As a result, it is self-evident that accurate defence must be developed. Intrusion detection approaches based on machine learning system (IDS) for the security of the system In general, an encroachment A detection system (IDS) is a programme or system that detects something.

Infectious activities and policy breaches in a network or system system. An IDS detects anomalies and inconsistencies. During the course of daily activities, behaviour on a network is observed. In a network or system that detects security threats or assaults

Denial-of-service attacks, for example, are a type of network security (Dos). An intruder the detecting system also aids in the location, decision, and control of objects. Unlawful system use, such as unauthorised access, or modification and annihilation there are several kinds of depending on the user's perspective intrusion detection systems.

LITERATURE SURVEY

- Neethu B. (2014) PCAA was provided for the Naive Bayes collection of features to develop a network intrusion detection system. The study utilised the KDD 1999 dataset for trials. When compared to neural network and tree algorithms techniques, the result demonstrates the effectiveness of the methodology, achieving a higher discovery rate, low time consumption, and low cost factor, with an accuracy of 94%.
- Naseer et al. (2018) proposed, constructed, and educated employing a variety of deep neural network frameworks, including RNNs, Autoencoders, and CNNs; their models were also trained and assessed using NSLKDD datasets. The DCNN and LSTM models performed with 85 percent and 89 percent accuracy, respectively.
- Zhang et al. (2017) proposed two types of network intrusion detection: direct and combination. Direct employs one method, while combination uses a combination of algorithms. Their proposal is for a novel directed acyclic graph (DAG) detection model based on belief rules (BRB). When compared to traditional detection models, the findings demonstrate that the DAG-BRB combinational model has a higher rate detection.
- Wang et al. (2017) proposed a hierarchically spatial system that learns network traffic aspects on its own. The system is an invasive detection system that learns spatial components and LSTM network characteristics using deep CNNs.
- Tiresias was proposed by Shen et al. (2018) as a system for predicting harmful actions using deep rooted learning. The system uses RNNs based on prior surveillance to estimate what will happen on a computer. The testing was carried out using a commercial IPS dataset. Even in a complicated situation, high precision and steady findings were maintained, indicating that the technique was effective in anticipating the future actions that may occur on a system with a correctness of 93%.
- Zhao et al. (2017) proposed a network attack identification archetypal that includes deep learning and flow computation, as well as an instantaneous detection and classification algorithm. Instantaneous detection may be achieved with flooding data processing, and taxonomy exactness can be improved using a technique. Several tests were carried out and comparisons were established using the CICIDS2017 dataset. When compared to traditional approaches, the results demonstrated a greater immediate detection effectiveness.

.

IMPLEMENTATION STUDY

1. Dataset Description

Because machine learning methods are used in applications for diverse network security tasks, extensive raw data are required to monitor network activities and discriminate between usual and suspicious traffic. Several efforts to generating network datasets have been carried out throughout the years. Most machine learning experiments have evaluated their work simulated or actual network data, just a few of the datasets that have been publically available, despite the fact that many of them are still private owing to security concerns. Despite the fact that various datasets have been created, actual IoT and network traffic statistics that cover novel Botnet situations are rare.

2. Feature Extraction

Flow based features were extracted from network traffic data using the CICFlowMeter [25]. It is a CIC-enabled network traffic flow generator that generates several network traffic features. It study the pcap file and develops a document with extracted features, as well as a csv file with the database. By extracting additional features of the data set, this method was primarily intended to increase the guessing skills of class dividers.

3. Data Preprocessing

Database is converted into a machine suitable for ML using pre-processing, data processing methods. This process also includes data cleaning, which involves removing external or suspicious data that may compromise the accuracy and efficiency of database.

4. Feature Selection

It is critical to reduce the number of features used in training and testing algorithms in order to build a lightweight security solution suitable for IoT systems [13]. As a feature selection strategy, we employed the Random Forest algorithm. It showed to be a useful tool for lowering dataset dimensionality. The model trains and responds faster when the input data features are reduced from over 80 network traffic features to seven. For the whole dataset, the relevance weights of the characteristics

5. Machine Learning Algorithms

All tests were performed on Python using Python ML packages. We divided the testing of machine learning algorithms into three categories: using the algorithms for each attack on database separately; apply algorithms throughout the database with features that include advanced features; and using algorithms throughout the database with a set of features that include advanced features.

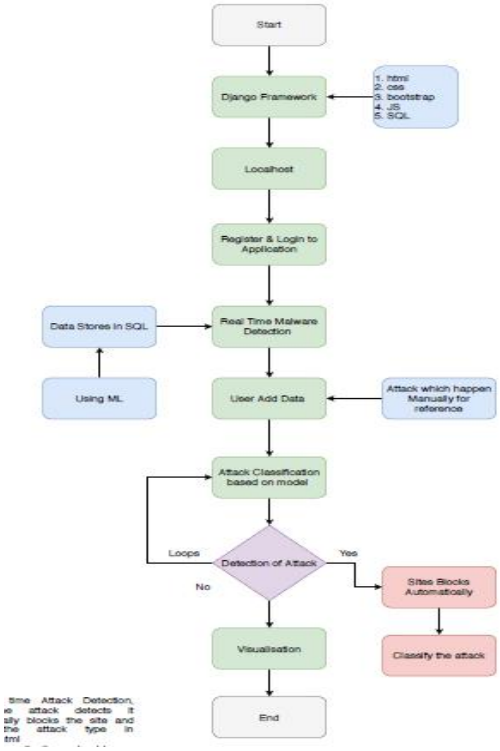

Fig 1: Proposed Approach

Proposed Approach

Important steps of the algorithm are given in below. 1) Normalization of every dataset. 2) Convert that dataset into the testing and training. 3) Form IDS models with the help of using RF, ANN and SVM algorithms. 4) Evaluate every model's performances.

1. Advantages

- Securing from malicious activities on network.
- Deleting and / or verifying malicious items within an existing network.
- Reject programs from certain potentially infected services.
- No access to confidential information.

Algorithms

1. Random Forest

Random Forest is a classifier that contains a no.of decision trees for the various data sets provided and takes measure to improve the prediction accuracy of that data.

Working of Random Forest Algorithm:

Step 1: First, start with selection of samples from given dataset.

Step 2: After that, this will create a decision tree for each sample. The result of the forecast on each decision tree will then be available.

Step 3: Each projected outcome will voted on this round.

Step 4: Lastly, select the predictable result with the most votes as the final prediction result.

2. ANN

MLP stands for multilayer perceptron and type of feedforward artificial neural network. Artificial neural networks are a type of machine learning method that is inspired by how the human brain learns and derives new information. An MLP has three layers. For training, MLP employs the unsupervised learning approach of back-propagation.

Results and Evolution Metrics

```
n [6]: columns=["duration","protocol_type","service","flag","src_bytes","dst_bytes","land",
       "wrong_fragment","urgent","hot","num_failed_logins","logged_in",
       "num_compromised","root_shell","su_attempted","num_root","num_file_creations",
       "num_shells","num_access_files","num_outbound_cmds","is_host_login",
       "is_guest_login","count","srv_count","serror_rate", "srv_serror_rate",
       "rerror_rate","srv_rerror_rate","same_srv_rate", "diff_srv_rate","srv_diff_host_rate","dst_host_count","dst_host_sr
       "dst_host_diff_srv_rate","dst_host_same_src_port_rate",
       "dst_host_srv_diff_host_rate","dst_host_serror_rate","dst_host_srv_serror_rate",
       "dst_host_rerror_rate","dst_host_srv_rerror_rate","attack", "last_flag"]

n [7]: train.columns=columns
       test.columns=columns

n [8]: train.head()

ut[8]:
```

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compromised	root_
0	0	udp	other	SF	146	0	0	0	0	0	0	0	0	
1	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	
2	0	tcp	http	SF	232	8153	0	0	0	0	0	1	0	
3	0	tcp	http	SF	199	420	0	0	0	0	0	1	0	
4	0	tcp	private	REJ	0	0	0	0	0	0	0	0	0	

```
n [9]: test.head()
```

Fig 2: Data Preprocessing and Extracted Features

```
# Protocol type distribution
plt.figure(figsize=(9,8))
sns.countplot(x="protocol_type", data=train)
plt.show()
```

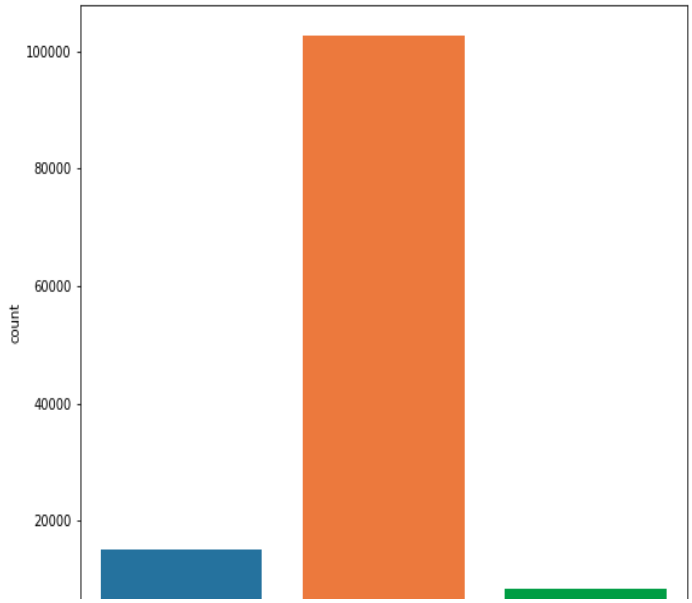

Fig 3: Fig count of records for training and Testing dataset

Table 1: Accuracy Result of the three different algorithms

ALOGITHAM	ACCURACY
Random Forest	99.82
ANN	99.70
SVM	93.29

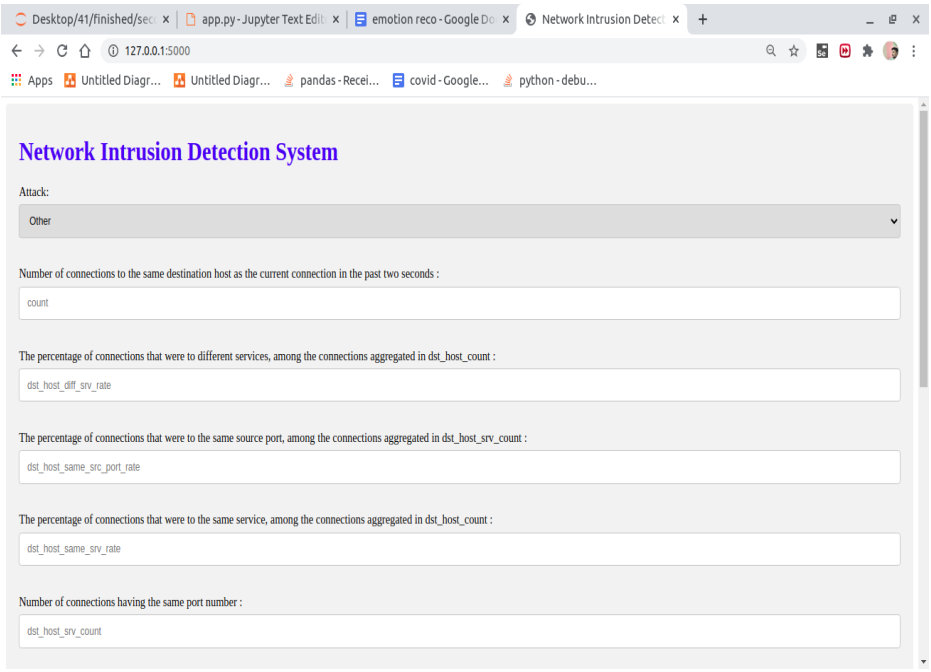

Fig 4: User Input Form for detection System

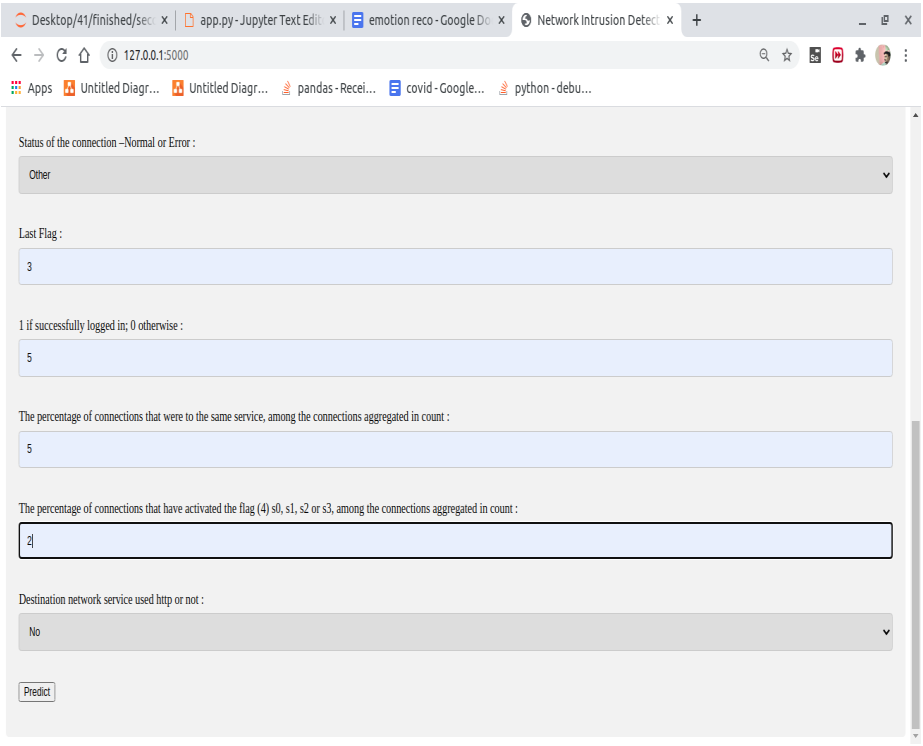

Fig 5: User Inputs of Different parameter values

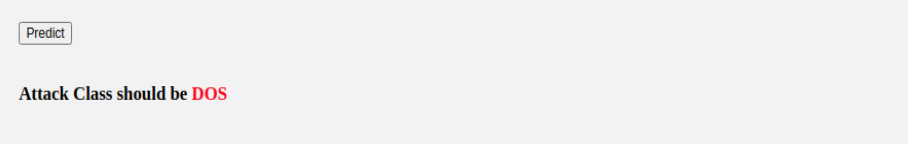

Fig 6: Predicted Attack Based on Feature Inputs

CONCLUSION

Estimates of assistance vector machine, artificial neural network, Voting classifiers, Random Forest, based on the contemporary CICIDS2017 dataset have been introduced very recently. The results reveal that the voting classifier calculation outperformed ANN, RF, in terms of performance. On the basis of this dataset, we use initiatives as well as other attack types with AI, Machine learning calculations. All of these calculations aid in the detection of a network cyber assault. When we go back over many years, there may have been a large number of attacks, and when these attacks are identified, the characteristics at which these attacks are occurring are saved in various datasets.

REFERENCES

Aiyanyo, I.D., Samuel, H., Heuiseok Lim, H. (2020). A Systematic Review of Defensive and Offensive Cyber security with Machine Learning. Appl. Sci. 2020, 10, 5811; doi: 10.3390/app10175811

Al-Haija, Q.A., Zein-Sabatto, S. (2020). An Efficient Deep-Learning-Based Detection and Classification System for Cyber-Attacks in IoT Communication Networks. Electronics 2020, 9, 2152; doi:10.3390/electronics9122152 www.mdpi.com/journal/electronics

Choi, H.Y., Sadollah, A., Kim, H.J. (2020). Improvement of Cyber-Attack Detection Accuracy From Urban Water Systems Using Extreme Learning Machine.

Chowdhury, S., Khanzadeh, M., et al. (2017). Botnet detection using graph-based feature clustering. Journal of Big Data, 4(1). https://doi.org/10.1186/s40537-017-0074-7

Ding, Y., Chen, S., Xu, J. (2016). Application of deep belief networks for opcode based malware detection. In Proceedings of 2016 International Joint Conference on Neural Networks (IJCNN), pp. 3901– 3908, Vancouver, British, July 2016.

Du, M., Li, F., Zheng, G., Srikumar, V. (2017). Deep Log: Anomaly detection and diagnosis from system logs through deep learning," in Proc. ACM CCS, Dallas, TX, USA, vol. 17, Nov. 2017, pp.1285_1298.

El-Alfy, E.M, Al-Obeidat, F. (2015). Detecting Cyber-Attacks on Wireless Mobile Networks Using Multi criterion Fuzzy Classifier with Genetic Attribute Selection. Hindawi Publishing Corporation. Mobile Information Systems. Volume 2015, Article ID 585432, 13 pages. http://dx.doi.org/10.1155/2015/585432

Fernández-Cabán, P.L., Masters, J.F., Phillips, B.M. (2018). Predicting Roof Pressures on a Low-Rise Structure from Freestream Turbulence Using Artificial Neural Networks. Frontiers in Built Environment. www.frontiersin.org 1 November 2018 | Volume 4 | Article 68.

Gershenson, C. (2003). Artificial Neural Networks for Beginners Graupe, D. (2007). Principles of Artificial Neural Networks (2nd Edition). Advanced Series on Circuits and Systems – Vol. 6. World Scientific Publishing Co. Pte. Ltd.

Gurney, K. (1997). An introduction to neural networks. University of Sheffield, UCL Press Limited Haddadi, F, C., Le, D. (2015). On the Effectiveness of Different Botnet Detection Approaches. Lecture Notes in Computer Science, 9065, 421–436. https://doi.org/10.1007/978-3-319-17533-1

Hodo, E., Bellekens, X., Hamilton, A., Dubouilh, P., Iorkyase, E., Tachtatzis, C., Atkinson, R. (2017). Threat analysis of IoT networks Using Artificial Neural Network Intrusion Detection System Hoque, S, A., Naser, A. (2012). An Implementation of Intrusion Detection System Using Genetic Algorithm. International Journal of Network Security & Its Applications, 4(2), 109–120.

https://doi.org/10.5121/ijnsa.2012.4208

Jacobson, L. (2013a). Introduction to Artificial Neural Networks - Part 1. https://www.theprojectspot.com/tutorial-post/introduction-to-artificial-neural-networks-part-1/7 JacobSon, L. (2013b). Introduction to Artificial Neural Networks Part 2 – Learning.

https://www.theprojectspot.com/tutorial-post/introduction-to-artificial-neural-networks-part-2- learning/8.

Khan, F.A., Gumaei, A., Derhab, A., Hussain, A. (2019). A novel two stage deep learning model for efficient network intrusion detection," IEEE Access, vol. 7, pp. 30373_30385, 2019.