ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Iss 12, 2022

Application of Machine Learning Techniques for the Detection of IoT Botnet Attacks

K. Sivanagireddy

Professor, Department of Electronics and Communication Engineering,
Sridevi Women's Engineering College, Hyderabad, India,
sivanagireddykalli@gmail.com

Abstract:

One of the primary cybersecurity threats faced by networking environments is Distributed Denial of Service (DDoS) attacks. As science and technology rapidly advance, organizations express growing concerns about their safety and security. DDoS attacks and computer infiltration attempts have consistently posed significant challenges in networked environments. DDoS attacks render services inaccessible to end-users by disrupting regular network flow and flooding the system with excessive packets, causing crashes. In the context of the Internet of Things (IoT), botnet attacks stand out as frequent and multi-stage cyber threats, initiating with scanning activities and culminating in distributed denial of service (DDoS) attacks. While existing research predominantly focuses on detecting botnet attacks after IoT devices are flooded with packets, our current approach employs machine learning algorithms such as logistic regression, SVM, random forest, and Gaussian distribution for proactive identification and prevention of botnet attacks.

1. Introduction

In recent years, security has grown to be a major concern in the internet world. Everything that is online has the potential to become vulnerable to several types of attacks. Botnet Attack is one of the internet attacks. It is a form of attack where the perpetrator gains access to the victim's system via bots. A Bot is a piece of application software that utilises the internet to execute automated operations or scripts. Simple and repetitive chores are carried out by these programmes much faster than by a human. A group of Internet-connected devices that are all running one or more bots is known as a botnet. Attacks like DDoS (Distributed Denial-of-Service) attacks and theft can be carried out using this. automated tasks or scripts over the internet. These applications perform tasks that are simple and repetitive but on much higher rate than any human.A Botnet is a collection of number of Internet-connected devices, with each one running one or more bots. This can be used to perform attacks like DDoS (Distributed Denial-of-service) attack, steal data, access to the device and its connection or send spam. The attacker can have control over the botnet using Command and Control Software. Let alone the risks of losing or stealing any data, the sheer use of botnet attack can cause harm to the victim devices. So, preventing such type of attack is not only important to various internet organizations but also to individuals. Organisations or Internet Service Providers that handle a lot of internet-related tasks frequently have protection to fend off these attacks. By examining the characteristics of the internet traffic and internet-related actions a device under inspection performs, UCI Machine several types of assaults can be detected. We will use the Learning Repository's IOT Botnet Detection dataset for this study.

2. Literature Review

[1].The Internet of Things (IoT), which uses methods to supply and confer information on the Internet, opens significant doors for wearable devices, home appliances, and programming. Given that specific normal data contains a lot of private information, maintaining information security on specific normal data is a significant concern that simply cannot be disregarded. In the aforementioned paper, we begin with a basic overview of IoT

ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Iss 12, 2022

information security before moving on to specific IoT information security-related issues. Finally, we will discuss research directions that are certain to be successful in the future while dealing with any implications relating to security issues that the IoT specifically faces.

[2] .Deep learning-based IoT botnet attack detection framework. The proposed solution specifically considered network traffic flows, which are further converted into feature records and then passed to the deep neural network (DNN) model for IoT botnet attack detection. Nugraha *et al.*evaluated the performance of four deep learning models for botnet attack detection by performing a couple of experiments. The experimental results revealed that CNN-LSTM outperformed all deep learning models for botnet attacks detection.

[3].Hybrid two-stage intrusion detection system (IDS) for the IoT environment. Their proposed approach first detects the anomalies from the network traffic, while in the second stage they classify the anomalies into attack classes. The authors used a multi-modal deep auto-encoder for anomalies detection, while used three machine learning classifiers to classify anomalies detected in the first stage. Likewise, Mirsky *et al* also used auto-encoders and proposed a plug and play network IDS, i.e., Kitsune to detect anomalies on local network traffic using an unsupervised learning approach. The authors used a self-generated botnet attack dataset and evaluated the performance in both online and offline modes. Their proposed solution achieved good performance comparable to offline anomaly detectors.

[4]. Hybrid botnet detection technique that detects new botnets implemented on three levels, i.e., host level, network level and a combination of both. The authors focused on focused HTTP, P2P, IRC, and DNS botnet traffic. The proposed technique consists of three components: host analyser, network analyser, and detection report. The authors used two machine learning algorithms, i.e., Naïve Bayes and a decision tree for traffic classification. Similarly, Blaise *et al.* proposed a bot detection technique named BotFP, for bot fingerprinting. The proposed BotFP framework has two variants, i.e., BotFP-Clus which groups similar traffic instances using clustering algorithms and BotFP-ML is designed to learn from the signatures and identify new bots using two supervised ML algorithms, i.e., SVM and MLP. Likewise, Soe *et al.* developed a machine learning-based IoT botnet attack detection model. The proposed model consists of two stages: a model builder and an attack detector. In the model builder stage, data collection, data categorization, model training and feature selection are performed step by step. While in the attack detector stage, the packets are first decoded and then the features are extracted in the same way as in the model builder phase. Finally, the features are passed to the attack detector engine where artificial neural network (ANN), J48 decision tree, and Naïve Bayes machine learning models are used for botnet attack detection.

3. Methodology

1. Scikit-learn: Scikit-learn is a popular machine learning library in Python that provides a wide range of algorithms for classification, regression, clustering, and more. It offers efficient implementations of algorithms like Decision Trees, Random Forests, SVM, Naive Bayes, and Logistic Regression, which can be used for IoT botnet attack detection.

2.Pandas: Pandas is a versatile data manipulation and analysis library in Python. It provides data structures and functions to handle datasets efficiently, such as loading data from CSV files, preprocessing data, performing feature engineering, and organizing data for training machine learning models. Pandas can be valuable for data preprocessing and preparation.

3.Numpy: Numpy is a fundamental numerical computing library in Python. It provides support for handling multi-dimensional arrays and matrices, along with mathematical functions and linear algebra operations. Numpy is commonly used in conjunction with other libraries for data manipulation and efficient computation during the training and evaluation stages.

4.Matplotlib : Matplotlib is a data visualization library in Python. This allows you

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 12, 2022

to create various types of plots, charts, and visual representations of data to analyze and present your results effectively. This library can be useful for visualizing the performance metrics, feature distributions, and any other relevant data during the detection process.

5. The sklearn.metrics module is part of the scikit-learn library, which is widely used for machine learning tasks in Python. The sklearn.metrics module provides various functions and classes for evaluating the performance of machine learning models. It offers a comprehensive set of metrics to assess the accuracy, precision, recall, F1-score, and other performance measures for classification, regression, and clustering tasks.

4. Implementation Algorithms Used

1. Logistic Regression: Logistic regression is a classification algorithm that is well-suited for binary classification problems. It estimates the probability of an instance belonging to a particular class. In the context of detecting IoT botnet attacks, logistic regression can be used to classify network traffic as either normal or malicious based on extracted features. It learns a linear decision boundary and uses a logistic function to model the probabilities.

2. Perceptron: The perceptron algorithm is a binary classification algorithm inspired by the functioning of a biological neuron. It learns a linear decision boundary to separate two classes. In the context of detecting IoT botnet attacks, the perceptron can be used to classify network traffic as normal or malicious based on extracted features. It updates its weights iteratively to minimize classification errors and converge towards a decision boundary.

3. Gaussian Naive Bayes: Naive Bayes is a probabilistic classification algorithm that assumes independence between features. Gaussian Naive Bayes is a variant of the algorithm that assumes a Gaussian (normal) distribution for the continuous features. In the context of detecting IoT botnet attacks, Gaussian Naive Bayes can be used to model the probability distribution of features in normal and malicious network traffic. It calculates the conditional probabilities of a new instance belonging to each class and selects the class with the highest probability.

4. Random Forest: Random Forest is an ensemble learning algorithm that combines multiple decision trees to make predictions. Each decision tree in the forest is trained on a random subset of the data and a random subset of features. In the context of detecting IoT botnet attacks, Random Forest can be used to classify network traffic based on a combination of features. It aggregates the predictions of individual trees to make a final prediction, improving accuracy and reducing overfitting.

5. KNN Classifier: KNN (k-nearest neighbors) is a non-parametric classification algorithm. It assigns a new instance to the class that is most common among its k nearest neighbors in the feature space. In the context of detecting IoT botnet attacks, KNN can be used to classify network traffic based on the similarity of feature vectors. The algorithm calculates the distances between instances and selects the k nearest neighbors to determine the class of the new instance.

5. Experimental Results

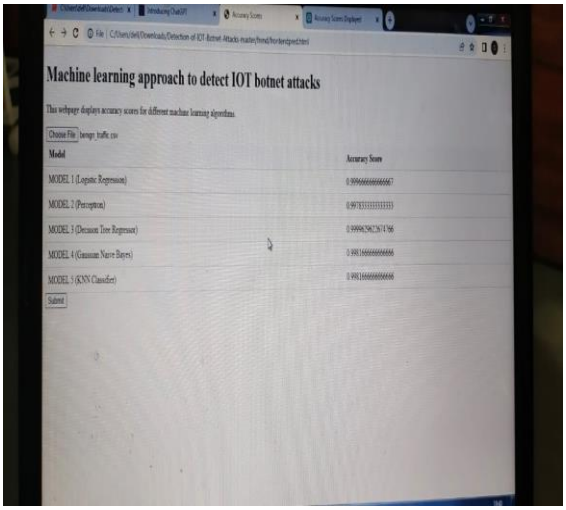

Fig1. Home Screen

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 12, 2022

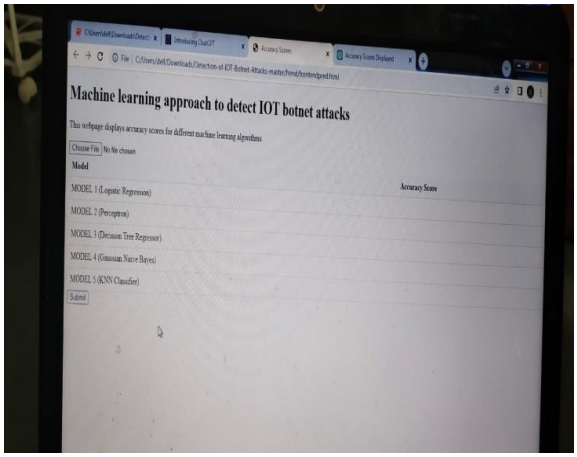

Fig2. After selecting dataset file the accuracy values

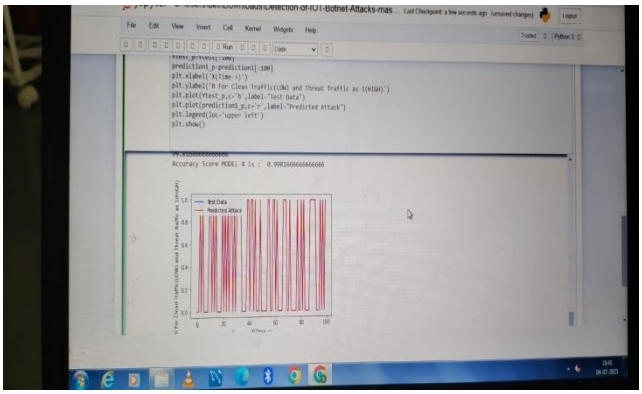

Fig3. Comparison Graph

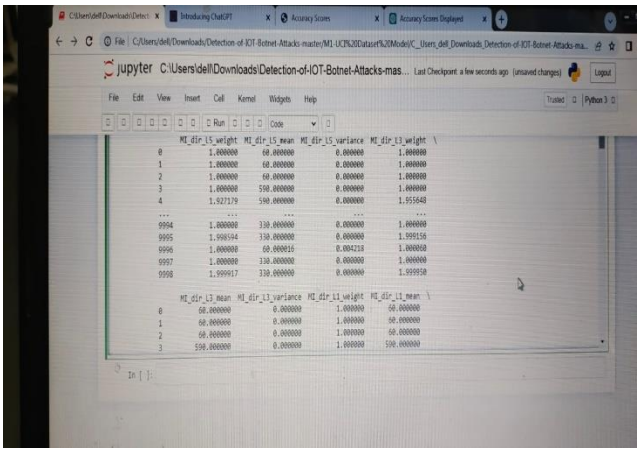

Fig4. Calculated mean, variance values

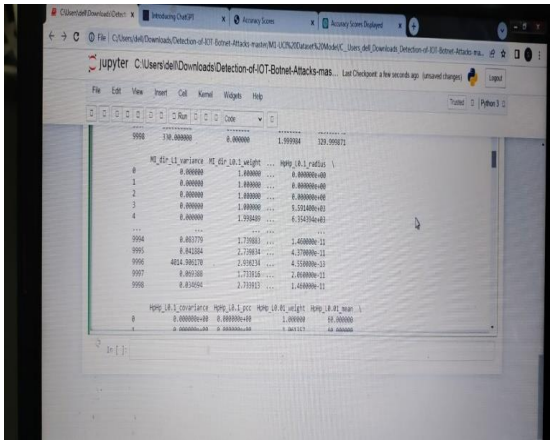

Fig5. Calculated covariance values

6. Conclusions & Future Scope

Model using UCI dataset had pretty ambiguous results. Though the dataset evidently illustrate that the difference in benign traffic and botnet traffic is clearly distinct, the accuracy of the models using UCI dataset is almost overwhelming. The accuracy comes in between 98% to 100% for some training models like Decision Tree, Logistic Regression, and even Perceptron. While other models can have varying accuracy between 80% to again 100%. To note, this is not a case of overfitting nor bad data, as different models found online also have same type of accuracy. Furthermore, even the data providers have mentioned study results concluding with 100 % True Positive Rate (TPR).

Integration with Network Security Systems: By combining the detection system with current network security infrastructure, the project can be expanded. Integration with firewalls, intrusion detection systems (IDS), and security information and event management (SIEM) systems are a few examples of this. Such integration can improve security posture overall and allow for a more thorough method of IoT botnet attack detection and prevention.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 12, 2022

References

1. Ali, A. I. A. Ahmed, A. Almogren, M. A. Raza, S. A. Shah, A. Khan, et al., "Systematic literature review on IoT-based botnet attack", *IEEE Access*, vol. 8, pp. 212220-212232, 2020.

2.S. Ghazanfar, F. Hussain, A. U. Rehman, U. U. Fayyaz, F. Shahzad and G. A. Shah, "IoT-Flock: An open-source framework for IoT traffic generation", *Proc. Int. Conf. Emerg. Trends Smart Technol. (ICETST)*, pp. 1-6, Mar. 2020.

3.M. Safaei Pour, A. Mangino, K. Friday, M. Rathbun, E. Bou-Harb, F. Iqbal, et al., "On data-driven curation learning and analysis for inferring evolving Internet-of-Things (IoT) botnets in the wild", *Comput. Secur.*, vol. 91, Apr. 2020.

4.F. Hussain, S. G. Abbas, M. Husnain, U. U. Fayyaz, F. Shahzad and G. A. Shah, "IoT DoS and DDoS attack detection using ResNet", *Proc. IEEE 23rd Int. Multitopic Conf. (INMIC)*, pp. 1-6, Nov. 2020.

5.S. Dange and M. Chatterjee, "IoT botnet: The largest threat to the IoT network" in Data Communication and Networks, Singapore:Springer, pp. 137-157, 2020.

6.F. Hussain, S. G. Abbas, U. U. Fayyaz, G. A. Shah, A. Toqeer and A. Ali, "Towards a universal features set for IoT botnet attacks detection", *Proc. IEEE 23rd Int. Multitopic Conf. (INMIC)*, pp. 1-6, Nov. 2020.

7.A. O. Prokofiev, Y. S. Smirnova and V. A. Surov, "A method to detect Internet of Things botnets", *Proc. IEEE Conf. Russian Young Res. Electr. Electron. Eng. (EIConRus)*, pp. 105-108, Jan. 2018.

8.B. K. Dedeturk and B. Akay, "Spam filtering using a logistic regression model trained by an artificial bee colony algorithm", *Appl. Soft Comput.*, vol. 91, Jun. 2020.

9.N. Vlajic and D. Zhou, "IoT as a land of opportunity for DDoS hackers", *Computer*, vol. 51, no. 7, pp. 26-34, 2018.

10.*GitHub Survived Biggest DDoS Attack Ever Recorded*, May 2021, [online] Available: https://github.blog/2018-03-01-ddos-incident-report/.

11.*AWS Said it Mitigated a 2.3 Tbps DDoS Attack Largest Ever*, May 2021, [online] Available: https://www.zdnet.com/article/aws-said-it-mitigated-a-2-3-tbps-ddos-attack-the-largest-ever/

12.*Shodan*, May 2021, [online] Available: https://www.shodan.io/.

13.*Censys*, May 2021, [online] Available: https://censys.io/.

14.C. Kolias, G. Kambourakis, A. Stavrou and J. Voas, "DDoS in the IoT: Mirai and other botnets", *Computer*, vol. 50, no. 7, pp. 80-84, 2017.

15.R. Hallman, J. Bryan, G. Palavicini, J. Divita and J. Romero-Mariona, "IoDDoS—The internet of distributed denial of sevice attacks", *Proc. 2nd Int. Conf. Internet Things Big Data Secur.*, pp. 47-58, 2017.

16.H.-T. Nguyen, Q.-D. Ngo and V.-H. Le, "A novel graph-based approach for IoT botnet detection", *Int. J. Inf. Secur.*, vol. 19, no. 5, pp. 567-577, Oct. 2020.

17.F. Hussain, S. G. Abbas, G. A. Shah, I. M. Pires, U. U. Fayyaz, F. Shahzad, et al., "A framework for malicious traffic detection in IoT healthcare environment", *Sensors*, vol. 21, no. 9, pp. 3025, Apr. 2021.

18.Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, et al., "N-BaIoT—Network-based detection of IoT botnet attacks using deep autoencoders", *IEEE Pervasive Comput.*, vol. 17, no. 3, pp. 12-22, Jul./Sep. 2018.

19.H. Hindy, D. Brosset, E. Bayne, A. Seeam, C. Tachtatzis, R. Atkinson, et al., "A taxonomy and survey of intrusion detection system design techniques network threats and datasets" in arXiv:1806.03517, 2018.