Paper Title\* (use style: paper title)

\*Note: Sub-titles are not captured in Xplore and should not be used

Madhuri Gurunathrao Desai   
 College of Computing and Software Engineering  
Kennesaw State UniversityAtlanta, USA  
mdesai9@students.kennesaw.edu

*Abstract*—Intrusion Detection is the art of detecting unauthorized, inappropriate, or anomalous activity on computer system. There are many devices connected in the network and in turn there are numerous activities taking place in the network. So, it is important to detect all kinds of intrusions in the system. Recently, many studies have demonstrated high accuracy of machine learning methods in detection the intrusion. However, requirement for classification and minimizing the feature is important to overcome computational problems. And, also detecting different types of attack helps the security analyst to attend and take actions on these attacks quickly. Our work focuses on detecting 3 types of attack which includes IoT botnet attack and other network attacks. To achieve this, we are building a multiclass classification using supervised learning models along with the Dimensionality Reduction. Numerous studies on ML-based IDS have been using KDD or the upgraded versions of KDD dataset. Here in this study we have used a new dataset, IoT network intrusion detection dataset.

Keywords—Supervised Learning, Multiclass Classification, Dimensionality Reduction, types of attacks, IoT network intrusion dataset

# Introduction

Internet of Things is one of the emerging technologies which is being used in various fields. These IoT devices are vulnerable to being weaponized with botnets for the purpose of carrying out different attacks like Mirai botnet attack, DDoS attack, Man-in-the-middle attack etc. Securing networks from intrusions or attacks is becoming harder as the network technologies are rapidly growing.

In this paper we have applied the standard workflow which consist of data preparation, feature reduction, model building, model training, validation, and result. Data preparation we have used a new dataset ‘IoT network intrusion detection dataset’. For feature reduction we have used PCA, SVD the reduced feature allows to minimize the computational speed. For the model building and training we have used Decision Tree, Random Forest, KNN, SVM, Neural Network, Naïve Bayes, logistic regression which are Supervised learning algorithms to detect the attacks. Here we have used multiclass classification to detect the attacks and categorize into four different types Benign, Mirai, Man-in-the middle and Scan attacks. These categorization helps the Security analyst to take the required actions on these attacks.

# Backgroung information

In this section we briefly discuss the background on the IoT botnet detection, network anomaly botnet detection, classifiers. Over the last decades many IoT botnet intrusions detection [1] and different types of attack in intrusion detections related papers have been proposed. Several machine learning (ML) algorithms, which includes both supervised and unsupervised Learning have been applied in IoT botnet detection, for instance DR for IoT botnet detection [1], in this paper they have classified the attacks into either benign or Mirai attack. They have used decision tree and F-score to build and train their model. There are also papers on unsupervised anomaly based botnet detection [5, 6], Neural Network [2], Decision Tree [3], botnet detection using Eigen space deep learning[7],and more have been extensively employed to detect intrusion activities from large quantity of complex and dynamic datasets. For the better computational speed, they have reduced the dimension using many dimensionality reduction algorithms.

All these papers have addressed the issue either related to IoT botnet intrusion or network intrusion detection. But when there is information being exchanged in the network with the connected devices then not only there is going to be a botnet attack but also, there are many different network intrusion (like Scan, protocol specific attacks etc.) so it is necessary to detect all these kinds of attacks and one of the possible ways is through multi class classification. There are few papers proposed on the multi class, in this paper they have categorized into two types normal and attack using the concept of deep learning [8], another paper where there have proposed the classification of the detailed information about 4 types of attack (DOS, U2R, R2L and Probing) [9], here they have used only one method SVM for the intrusion detection using KDD cup dataset.

Of all this there is another important issue which needs to be addressed which is the attack detection decision. Some of the papers address the issue with only the normal and mirai detection whereas if all the attacks present in the system are detected then it will be easy for the security analyst to take actions on the attack quickly.

# Methods

We have used python coding for the model building using below methods.

## Decision Tree

Step 1: creating decision tree:

In this method the dataset is split into training (80%) and testing (20%) data. Here the labels are been created to classify into normal and different attacks. We have used the random.seed() method to generate the random numbers. A decision tree classifier method is called through the python package, here we have kept the decision tree depth as, maximum *depth=3* and *criterion is set to ‘entropy’.* We have defined different levels of depth as 3,5,7,9,11. Then we take the root mean square error (RMSE) of both training and testing data. Then the accuracy of the model is calculated.

First, we have passed the 115 features extracted 36 files (which includes all the attacks and normal traffic) to the decision tree model we have built. The statistics we have used to measure the performance of the models is accuracy and RMSE (Root Mean Square Error). The RMSE is calculated using the formula below

The accuracy obtained I for the Decision tree is training accuracy-98.3% and testing accuracy-98.2%. the final training RMSE- 0.15 and final testing RMSE is 0.20.

|  |  |  |  |
| --- | --- | --- | --- |
| **Methods** | **DR- Methods** | **Accuracy in %** | **RMSE** |
|  |  | **Train Test** | **Train Test** |
| Decision  Tree | SVM | 98.3 98.2 | 0.15 0.20 |
|  | PCA |  |  |
| Random Forest | SVM |  |  |
|  | PCA |  |  |

**TABLE 1: Accuracy and RMSE values**

# Experiments

## Dataset

## The KDD CUP 1999 dataset (KDD) is developed by Defense Advanced Research Projects Agency (DARPA) and is the most used dataset for IDS evaluation [3]. This dataset is being used in almost all intrusion detection papers. The KDD classifies attacks into four categories, such as DoS, User to Root (U2R), Remote to Local (R2L) and Probing. KDD was generated by injecting these kinds of attacks into each category. Numerous IDS studies have been using KDD as a dataset since machine learning is actively employed into IDS studies. Most of these studies perform binary classification that classifies the entire KDD into attack and benign. They also carry out multiclass classification to classify the KDD into the four categories Maintaining the Integrity of the Specifications.

In this study we have used a new dataset ‘IoT network intrusion detection dataset’[10]. As per my knowledge this dataset is not used in any intrusion detection papers. This dataset is from HCRL (Hacking and counter measure research lab) and was added on September 20, 2019. To download the dataset, we had submitted the application form with all the required details like purpose for download, email id etc. The link to the download was sent to the email id where we were able to download the required dataset. This dataset is created with various types of network attacks in Internet of Things (IoT) environment. Two smart home devices (SKT NUGU (NU 100) and EZVIZ Wi-Fi Camera (C2C Mini O Plus 1080P)) were used, including laptops and smartphones which are connected to same wireless network. The dataset consists of 36 raw network packet files (pcap) at different time points. All attacks except Mirai Botnet category are the packets captured while simulating attacks using Nmap tool. In case of Mirai Botnet category, the attack packets were generated on a laptop and then manipulated to make it appear as if it originated from the IoT device. IoT network intrusion detection dataset is categorized into 3 different types of attacks Mitm (Man in the middle), Mirai and Scan. The dataset has normal (1,756,276), man-in the middle (101,885), Scan (25210), Mirai (987977).

## Feature Extraction

We have extracted 115 features from each record using feature extraction method [11]. The downloaded dataset were pcap(wire shark) readable files which are the raw data files, and these had to be converted to (.csv) readable files. This is applied to the above described dataset ‘IoT network intrusion detection dataset’. First, we analyzed what are the IP-address of each of the target devices. Later, the different analysis was made like, In this method recent history of the stream is captured in these time statistics L5(100ms), L3(500ms), L1(1.5sec), L0.1(10sec) and L0.01(1min). The traffic summary is: it is divided into five feature categories, Source-IP(H), Source-MAC&IP(MI), channel (HH), channel-jitter (HH\_Jit) and socket (HpHp). The detailed information is shown in TABLE 2. In this paper how each of the feature is extracted from category is for ex: Host IP-100ms- weight(Pkt Count)” corresponds to the feature that is computed by the packet count of the Source-IP category at interval 100ms. This is carried out for all the 5-time intervals to the feature stats of the particular feature category.

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Categories** | **Description** | **Extracted features Stat** | **No: of features from each category** |
| Source-IP(H) | Stats summarizing the recent traffic from pkt Source (IP) | **Weight** (pkt count), **Mean**,  **Variance** | **15** |
| Source-MAC&IP(MI) | Stats summarizing the recent traffic from pkt Source (IP + MAC) | **Weight** (pkt count), **Mean,**  **Variance** | **15** |
| Channel (HH) | Stats summarizing the recent traffic going from this pkt source (IP) to pkt destination. | **Weight** (pkt count), **Mean,**  **Std,**  **Magnitude** (root sq. sum of the two streams), **Radius** (variances),  **co-variance** (cov- appr. covariance between two streams),  **Pcc** (approx. correlation coefficient between two streams) | **35** |
| Channel-jitter (HH\_Jit | Stats summarizing the jitter of the traffic going from this pkt source (IP) to pkt destination. | **Weight** (pkt count), **Mean**,  **Variance** | **15** |
| Socket (HpHp) | Stats summarizing the recent traffic going from this pkt source+port (IP) to the pkt destination source+port | **Weight** (pkt count), **Mean,**  **Std,**  **Magnitude** (root sq. sum of the two streams), **Radius** (variances),  **co-variance** (cov- appr. covariance between two streams),  **Pcc** (approx. correlation coefficient between two streams) | **35** |

**TABLE 2: Attribute and Features information**

##### References

1. Hayretdin Bahs i, Sven N˜omm, Fabio Benedetto La Torre “Dimensionality Reduction for Machine Learning Based IoT Botnet Detection”,2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV) Singapore, November 18-21, 2018
2. Pal and M. A. M. Hasan, “Neural Network & Genetic Algorithm Based Approach to Network Intrusion Detec-tion & Comparative Analysis of Performance,” Proceedings of the the 15th International Conference on Computer and Information Technology, Chittagong, Bangladesh,2012.
3. <https://kdd.ics.uci.edu/databases/kddcup99/task>.
4. Shengchu Zhao1, Wei Li1, Tanveer Zia2 and Albert Y. Zomaya1, “A dimensional reduction model and classifier
5. Hayretdin Bahsi, Sven N˜omm “Unsupervised anamoly based botnet detection in IoT network”, 2018 17th IEEE International Conference on Machine Learning and Applications
6. “A Dimension Reduction Model and Classifier for Anomaly-Based Intrusion Detection in Internet of Things”
7. “Robust Malware Detection for Internet of (Battlefield) Things Devices Using Deep Eigenspace Learning”
8. Hector Alaiz-Moreton, Jose Aveleira-Mata, Jorge Ondicol-Garcia, Angel Luis Muñoz-Castañeda, Isaías García,and Carmen Benavides, “Multiclass Classification Procedure for Detecting Attacks on MQTT-IoT Protocol”,
9. Hansung Lee, Jiyoung Song, and Daihee Park, “Intrusion Detection System Based on Multi-class SVM”
10. “<http://ocslab.hksecurity.net/Datasets/iot-network-intrusion-dataset>”
11. Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, D. Breitenbacher,A. Shabtai, and Y. Elovici, “N-baiot: Network-based detection of iotbotnet attacks using deep”