Intrusion Prevention System

SHUBHAM SINGH
IT Department
IIITA
Prayagraj,India
mit2021023@iiita.ac.in

***Abstract*—Intrusion prevention system in IoT devices are the procedures that are treated as Add-ons' of the intrusion detection system to actively defend and prevent the intrusions, that are detected by the detection procedures of the IDS. Analysis has been made on two datasets UNSW-NB15 and TON-IOT20 by applying various machine learning methods for classification. For Intrusion Prevention System, two level classifier for dropping of packets. Level 1 classifier comprises of Decision Tree whereas Level 2 classifier uses Random Forest. Intrusion Prevention system which automatically adapts to system changes and further reduces training time and increases the accuracy of the system.**

I. Introduction

The most important layer of protection for IoT is undoubtedly securing the data, network and communication among IoT devices and with the Internet.

Challenges of Iot IoT security is challenged by constrained resources, in particular in what respects memory and energy, limited computational power, the usage of insecure operating system, insufficient authentication and authorization, lack of transport encryption, insecure network access and weak interfaces, among others[1].

Attack In the year 2016, the famous Mirai Botnet targeted IoT devices such as IP Cameras, baby monitors, printers, home routers and gateways and, in consequence, several Internet services in North America and Europe were brought down. Certainly there are two lines of defences[1].

A. *First line of defence*

This protection layer includes firewalls and cryptographic primitives of authentication, encryption, access control and secure key management, among others.

B. *Second line of defence*

IDS is any hardware or software that identifies and detects intrusions. An extension of IDS are Intrusion Prevention Systems (IPS), often referred to as inline IDS, and such systems aim to detect and prevent intrusions in real time. However, both IDS and IPS are useless without Intrusion Response Systems (IRS), which implement security countermeasures to monitor system performance and even identify and handle potential intrusions. IDS, IPS and IRS systems can collectively be termed as defensive security systems for intrusions.

C. *Overview*

1) IPS: IPSs, on the other hand, are considered extensions of IDS because they monitor system or network activity and attempt to stop/block intrusions. Unlike IDS, IPS are placed in-line and are able to proactively prevent intrusions that are detected. More precisely, IPS can take actions such as dropping malicious packets, sending alarms, resetting the connection, correcting transmission errors, cleaning unwanted network and transport layer options etc.

2) IDS: A typical IDS includes three generic components, namely monitoring, detection and reaction. Monitoring component analyse the behaviour of traffic flows. The detection component detects any suspicious behaviour and informs the reaction component of any detected occurrence. The reaction component raises an alarm or reports to the network administrator.

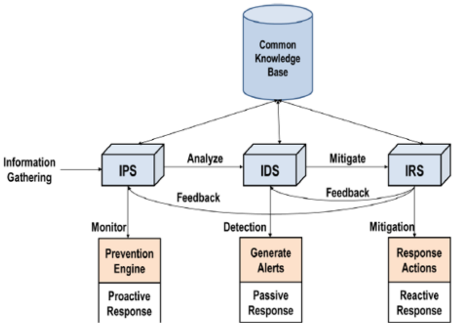

Fig. 1: IPS, IDS, IRS[1]

II. Literature Review

Intrusion Prevention System for DDoS attack on VANET[4] with reCAPTCHA Controller using information-based metrics published on 4 Oct 19. The reCAPTCHA controller mechanism in this method prohibits bots from launching consistent automated attacks, as well as many automated DDoS attacks. This method examines the randomness of user requests using information theory-based measures, such as entropy. Both frequency and entropy are utilised to determine the attack's vulnerability. DDoS attacks are carried out using AODV, Firecol, and reCAPTCHA. It has been proven that reCAPTCHA performs well during DDoS attacks.

Intrusion prevention system focused on big data[5] is published on 10 May 2021. This research presented a huge data-based hierarchical deep learning system to improve efficiency. Captures both network traffic and content information using behavioural and content-functional capability. This model enhances the accuracy of fake attack categorization. SVM and DT are the algorithms employed. The independent test and the 5 2 cross-validation F test are used to assess IDS performance. The function extraction and selection method may consume too much storage space for large datasets. Storage for sparks For an efficient approach to role collecting, the MapReduce architecture is used.

Network Intrusion Prevention System in Real-Time[6] According to Hybrid Machine Learning, which was released on March 17, 2021, it is capable of high performance and real-time categorization. This system employs two classifiers. Only if the Level 1 classifier can categorise packets as normal or attack with precision can it receive them at line speed. Level 2 classifier has no real-time processing limits, therefore it can take any leftover packets that were not classified as attack in level 1 and categorise them slowly but accurately. The dataset is the same in both.

Enhancing Cyber Security in Modern Internet of things (IoT) Using Intrusion Prevention Algorithm for IoT (IPAI)[7] published in 2019. The suggested technique focuses on the investigation and research of bandwidth attacks, with a special focus on DDoS, which is a difficult-to-detect problem that degrades network performance. DDoS is a sort of assault in which a collection of attacker nodes attacks the target, preventing legitimate users from accessing network services and resources. The proposed technique is based on the report generated by the IDS after analysing the forensic analysis report, which further leads to the prevention of DDoS assaults in the network.

III. Problem statement

Analysis of UNSW-NB15[8] and TON-IOT (UNSW-IOT20)[9] datasets by applying various machine learning algorithms for Detection Phase.

IV. Datasets

A. UNSW-NB15[8]

UNSW-NB 15 data set is created by the IXIA PerfectStorm tool in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) for generating a hybrid of real modern normal activities and synthetic contemporary attack behaviours. This data set has nine families of attacks, namely, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms. This dataset consists of 49 features, and are described in UNSW-NB15_features.csv file. Dataset used for binary and multiclass classification is UNSW_NB15_training-set.csv. The number of records in the Dataset is 175,341 records from different the types of attack and normal.

Dataset Features

- **Nominal :** 'proto', 'service', 'state', 'attack_cat'
- **Integer :** 'sbytes', 'dbytes', 'sttl', 'dttl', 'sloss', 'dloss', 'swin', 'stcpb','dtcpb', 'dwin', 'trans_depth', 'ct_srv_src', 'ct_state_ttl', 'ct_dst_ltm', 'ct_src_dport_ltm', 'ct_dst_sport_ltm', 'ct_dst_src_ltm', 'ct_ftp_cmd', 'ct_flw_http_mthd', 'ct_srv_dst'
- **Float :** 'dur', 'tcprtt', 'synack', 'ackdat'
- **binary :** 'is_ftp_login', 'is_sm_ips_ports'

For Binary Classification, feature label has two labels as 19,488 records as normal data and 61685 records as attack data. For Multiclass Classification, feature attack_cat has 9 labels Analysis(564), Backdoor(99), Dos(1,791), Exploits(16,187), Fuzzers Multi labels(1,731), Generic(39,496), Normal(19,488), Reconnaissance(1,703), Worms(114).

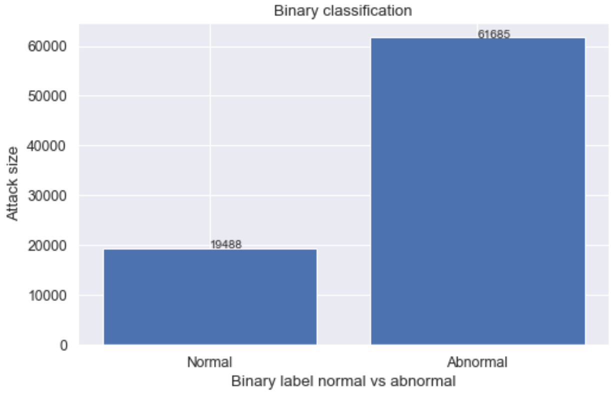

Fig. 2: Binary classification (UNSW-NB15)[10]

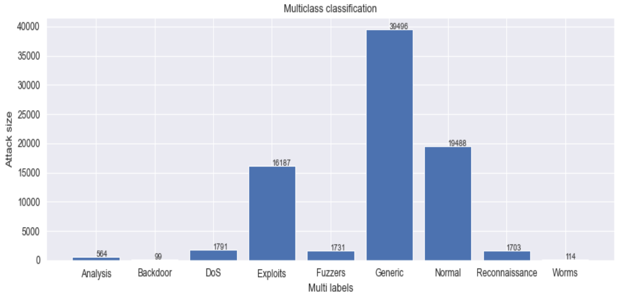

Fig. 3: Multiclass classification(UNSW-NB15)[10]

B. TON_IoT (UNSW-IoT20)[9]

The TON_IoT (UNSW-IoT20) datasets are new generations of Internet of Things (IoT) and Industrial IoT (IIoT) datasets for evaluating the fidelity and efficiency of different cybersecurity applications based on Artificial Intelligence (AI). The datasets have been called 'ToN_IoT' as they include heterogeneous data sources collected from Telemetry datasets of IoT and IIoT sensors, Operating systems datasets of Windows 7 and 10 as well as Ubuntu 14 and 18 TLS and Network traffic datasets. The datasets were collected from a realistic and large-scale network designed at the IoT Lab of the UNSW Canberra Cyber, the School of Engineering and Information technology (SEIT), UNSW Canberra @ the Australian Defence Force Academy (ADFA). The datasets were gathered in a parallel processing to collect several normal and cyber-attack events from IoT networks.

Processed IoT dataset is being used for binary and multiclass classification. It consists of seven .csv files IoT Fridge.csv, IoT Garage_Door.csv, IoT GPS_Tracker.csv, IoT Modbus.csv, IoT Motion_Light.csv, IoT Thermostat.csv, IoT Weather.csv .

Dataset along with features

- **IoT Fridge:** ts, date, time, fridge temperature, temp condition, label, type
- **IoT Garage_Door:** date, time, door state, sphone signal, label, type
- **IoT GPS_Tracker:** date, time, latitude , longitude, label, type
- **IoT Modbus:** date, time, FC1 Read Input Register, FC2 Read Discrete Value, FC3 Read Holding Register, FC4 Read Coil, label, type
- **IoT Motion_Light :** date, time, motion status, light status, label, type
- **IoT Thermostat :** date, time, current temperature, thermostat status, label, type
- **IoT Weather :** date, time, temperature, pressure, humidity, label, type

Each dataset alongside with attack label's percentage

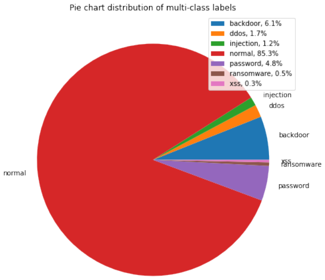

Fig. 4: IoT Fridge dataset[10]

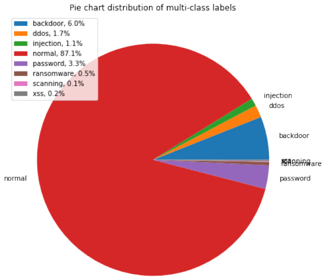

Fig. 5: IoT Garage Door dataset[10]

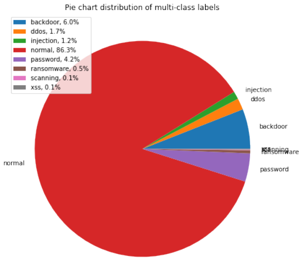

Fig. 6: IoT GPS Tracker dataset[10]

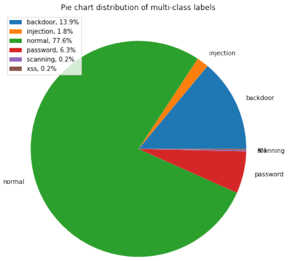

Fig. 7: IoT Modbus dataset[10]

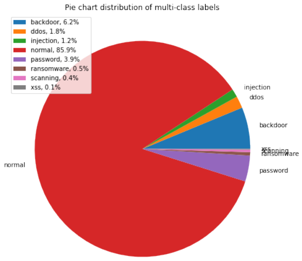

Fig. 8: IoT Motion Light dataset[10]

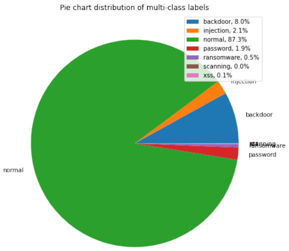

Fig. 9: IoT Thermostat dataset[10]

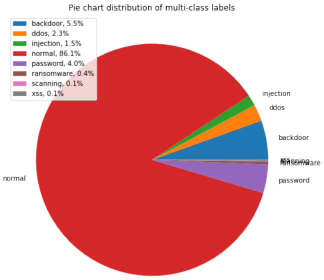

Fig. 10: IoT Weather dataset[10]

V. Methodology

A. UNSW-NB15[8]

1) Data Pre-processing and Normalization: Dataset's feature 'select' consists of values '-' so entire rows are deleted from dataset. Variant numeric data types are converted into single numeric data type.
Nominal/categorical data is dealt using **one-hot encoding** i.e., features which lie in this category are **proto**, **service**, **state**.
Total features after encoding are 61. All numeric datatype features are normalized using **MinMaxScalar()** with range(0,1).
Binary labels are formed using **LabelEncoder()**, where as multiclass labels are formed using **one-hot-encoding** and **LabelEncoding**.

2) Feature Selection: Correlation matrix is formed and features with correlation value less than 0.3 are removed from dataset. This strategy is applied for binary data and multiclass data selection.

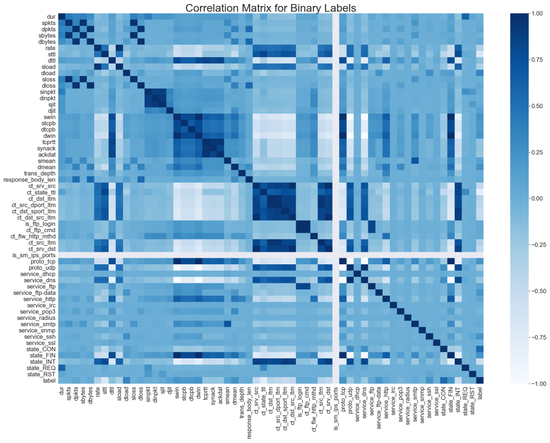

Fig. 11: Correlation matrix Binary Data (UNSW-NB15)[10]

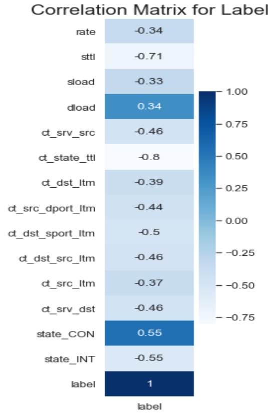

Fig. 12: Feature Selection Binary Data (UNSW-NB15)[10]

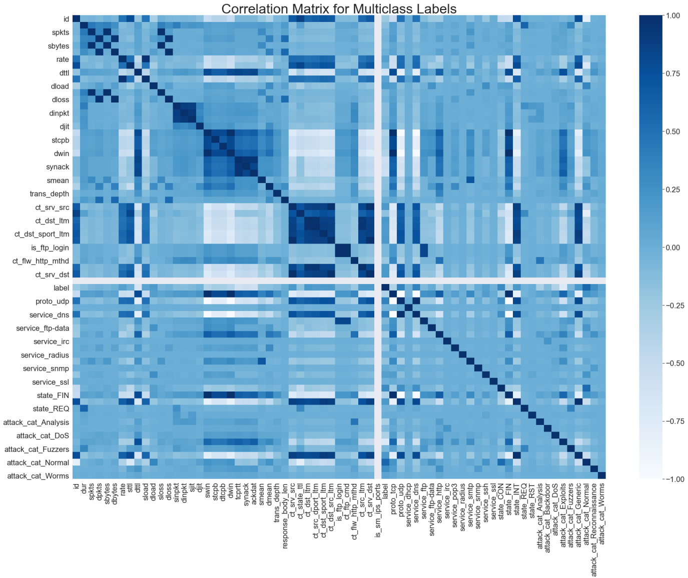

Fig. 13: Correlation matrix Multiclass data (UNSW-NB15)[10]

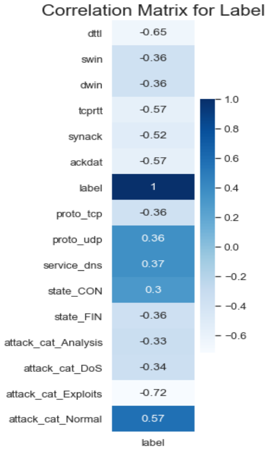

Fig. 14: Feature Selection Multiclass Data (UNSW-NB15)[10]

B. TON-IOT (UNSW-IOT20)[9]

1) Data Pre-processing: From all seven datasets, **timestamp** , **data** , **time** has been removed.

In dataset **IoT Fridge** , feature **temp_condition** has six unique values ‘high’, ‘high ‘, ‘high ‘, ’low’, ‘low ‘, ‘low ‘ and further processed by trimming down the space making it **high** and **low**.
In dataset IoT Garage Door, feature **sphone_signal** consists of six unique labels ‘0’, ’false ‘, ‘0.0’, ‘1’, ‘true ‘, ‘1.0’ and further processed by omitting it to **false** and **true**.

2) Normalization: All numeric data type features are normalized using **MinMaxScaler(MinMaxScaler()** with range(0,1).
In Dataset IoT Fridge, feature **fridge_temperature** has been normalized.
In Dataset IoT GPS Tracker, feature **latitude** and **longitude** has been normalized.
In Dataset IoT Modbus, features **FC1_Read_Input_Register**, **FC2_Read_Discrete_Value**, **FC3_Read_Holding_Register**, **FC4_Read_Coil** has been normalized.
In Dataset IoT Thermostat, feature **current_temperature** has been normalized. In Dataset IoT Weather, feature **temperature**, **pressure** and **humidity** has been normalized.

VI. Models used for analysis

- **Logistic Regression**
- **Naïve Bayes**
- **KNN**
- **Decision Tree**
- **Random Forest**
- **AdaBoost**
- **SVM linear, rbf, sigmoid**

VII. Observation

A. Selection of Models

Binary classification and Multiclass classification on UNSW-NB15 is depicted by Tables

TABLE I: Binary classification - UNSW-NB15[9]

ML Model	Acc.	Prec.	Recall	F1
LR	0.977	0.98	0.98	0.98
NB	0.745	0.87	0.75	0.76
KNN	0.982	0.98	0.98	0.98
DT	0.980	0.98	0.98	0.98
RF	0.984	0.98	0.98	0.98
AdaB	0.981	0.98	0.98	0.98
SVM	0.966	0.96	0.96	0.96

TABLE II: Multiclass classification - UNSW-NB15[9]

ML Model	Accuracy	Precision	Recall	F1
LR	0.974	0.97	0.97	0.97
NB	0.951	0.95	0.95	0.95
KNN	0.972	0.97	0.97	0.97
DT	0.970	0.97	0.97	0.97
RF	0.973	0.97	0.97	0.97
AdaB	0.751	0.63	0.75	0.67
SVM	0.973	0.97	0.97	0.97

So for Intrusion Prevention System Models to be used are Decision Tree Random Forest.

B. Intrusion Prevention System[6]

1) Phase 1 : Formation of models and calculation of threshold value: firstly, both models are trained using Decision Tree Random Forest. Threshold value θ is calculated by find the average time taken for classifying the training dataset.

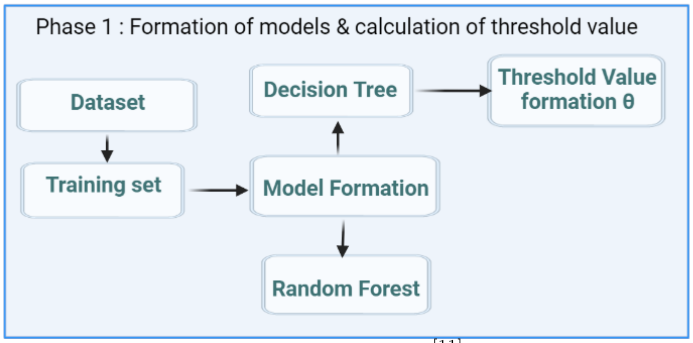
Phase 1 : Formation of models & calculation of threshold value
Dataset
Training set
Decision Tree
Model Formation
Random Forest
Threshold Value formation θ

Fig. 15: Phase 1 : IPS[11]

2) Phase 2 : Prevention - Dropping of packets: Now, Decision Tree Model is used for Level 1 classifier whereas Random Forest is used for Level 2 classifier. If the value is less than or equal to threshold value θ than L1 classifier is used otherwise stream is forward to L2 classifier. Both L1classifier , L2 classifier are used for prevention of attack.

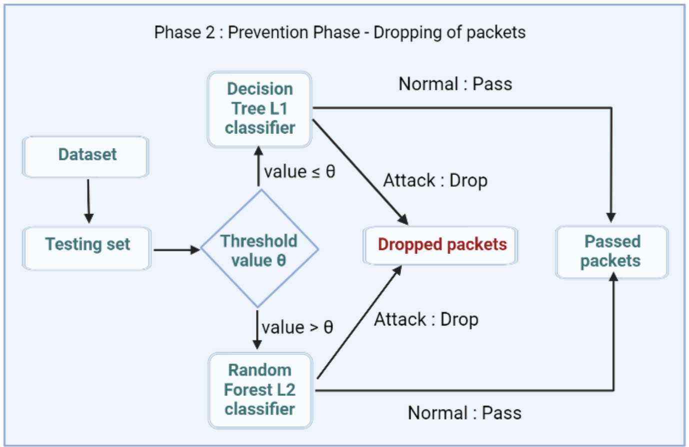

Fig. 16: Phase 2 : IPS[11]

Furthermore performance metrics for both binary and multi-class classifiers are:

TABLE III: Binary classifiers

ML Model	Accuracy	Precision	Recall	F1
DT(L1)	0.981	0.98	0.98	0.98
RF(L2)	0.982	0.98	0.98	0.98

TABLE IV: Multi-class classifiers

ML Model	Accuracy	Precision	Recall	F1
DT(L1)	0.970	0.97	0.97	0.97
RF(L2)	0.972	0.97	0.97	0.97

VIII. Conclusion

Prevention techniques may involve packet drop which leads to data loss. Higher accuracy models require more computational memory which makes method less real-time prevention and detection. Whereas signature-based methods have low computational cost and low accuracy which is again not ideal in real-world[6].

References

[1] Kamaldeep, M. Dutta and J. Granjal, ”Towards a Secure Internet of Things: A Comprehensive Study of Second Line Defense Mechanisms,” in IEEE Access, vol. 8, pp. 127272-127312, 2020

[2] V. Paxson. (2018). The Zeek Network Security Monitor. [Online]. Available: https://www.zeek.org

[3] S. Anwar, J. Mohamad Zain, M. F. Zolkipli, Z. Inayat, S. Khan, B. Anthony, and V. Chang, “From intrusion detection to an intrusion response system: Fundamentals, requirements, and future directions,” Algorithms, vol. 10, no. 2, p. 39, Mar. 2017.

[4] M. Poongodi, V. Vijayakumar, F. Al-Turjman, M. Hamdi and M. Ma, ”Intrusion Prevention System for DDoS Attack on VANET With reCAPTCHA Controller Using Information Based Metrics,” in IEEE Access, vol. 7, pp. 158481-158491, 2019, doi: 10.1109/ACCESS.2019.2945682.

[5] Alqahtani, A.S., Abuhasel, K.A. Alquraish, M. On implementing a powerful intrusion prevention system focused on big data. J Supercomput 77, 14039–14052 (2021). https://doi.org/10.1007/s11227-021-03856-8

[6] W. Seo and W. Pak, ”Real-Time Network Intrusion Prevention System Based on Hybrid Machine Learning,” in IEEE Access, vol. 9, pp. 46386-46397, 2021, doi: 10.1109/ACCESS.2021.3066620.

[7] A. Aldaej, ”Enhancing Cyber Security in Modern Internet of things (IoT) Using Intrusion Prevention Algorithm for IoT (IPAI),” in IEEE Access, doi: 10.1109/ACCESS.2019.2893445.

[8] https://research.unsw.edu.au/projects/unsw-nb15-dataset

[9] https://research.unsw.edu.au/projects/toniot-datasets

[10] https://matplotlib.org/

[11] Created with BioRender.com