Intrusion Detection System in IoT using Machine Learning

B23BH03 | BTP

B Kartheek (S20200010030)

M Sathwic (S20200010114)

Kush Gupta (S20200010108)

Dr.Bheemappa Halavar

Overview

Introduction

Motivation

Problem Statement

Why UNSW-NB15 dataset ?

Exploratory Data Analysis

Plan of Work

Model Implementation

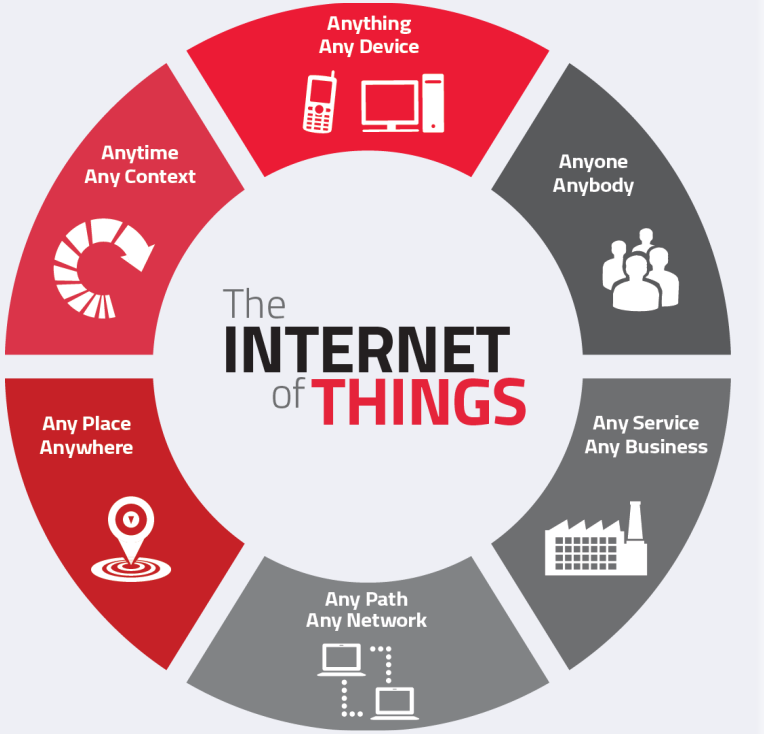
Future Scope of Project

References

Introduction

What is an IoT Device?

- An IoT device refers to any physical device that can be connected to the internet and can communicate with other devices or systems over the internet.
- The IoT devices include wireless sensors, software, actuators, computer devices and more.



Motivation

Growing Threats to Cybersecurity

Cyber attacks are increasing in sophistication and frequency. Traditional intrusion detection systems are limited in their ability to detect new threats.

Advancements in Machine Learning

Machine Learning algorithms have shown great potential in intrusion detection. They can learn from historical data and adapt to changing environments.

Importance of Intrusion Detection

Intrusion detection is critical for preventing unauthorized access and protecting sensitive information. It can save organizations time and money.

Relevance to Computer Science

This BTP project will give us experience developing a real-world application using Machine Learning algorithms.

Problem Statement

- **Due to restrictions such as low processing power and limited memory, traditional intrusion detection systems (IDSs) may be ineffective in detecting new and sophisticated assaults targeting IoT devices.**
- **An efficient and effective IDS for IoT that can identify aberrant behaviour in real-time and prevent possible security breaches is required.**
- **Machine learning (ML) has showed promise in improving the detection accuracy of intrusion detection systems (IDSs), but its application to IoT devices remains difficult.**

Why UNSW-NB15 dataset ?

The UNSW-NB15 dataset is a network security dataset that contains over 2 million instances of normal and attack traffic.

1 The dataset is realistic and diverse

UNSW-NB15 captures the complexities and subtleties of modern network attacks, making it a top choice among other datasets.

2 Includes labeled and unlabeled data

UNSW-NB15 covers various network protocols and services such as HTTP, DNS, FTP, and SMTP, providing a more comprehensive dataset than other publicly available options.

3 Unique types of attacks

UNSW-NB15 includes several types of attacks that are not commonly found in other datasets, including SQL injection and HTTP flood attacks.

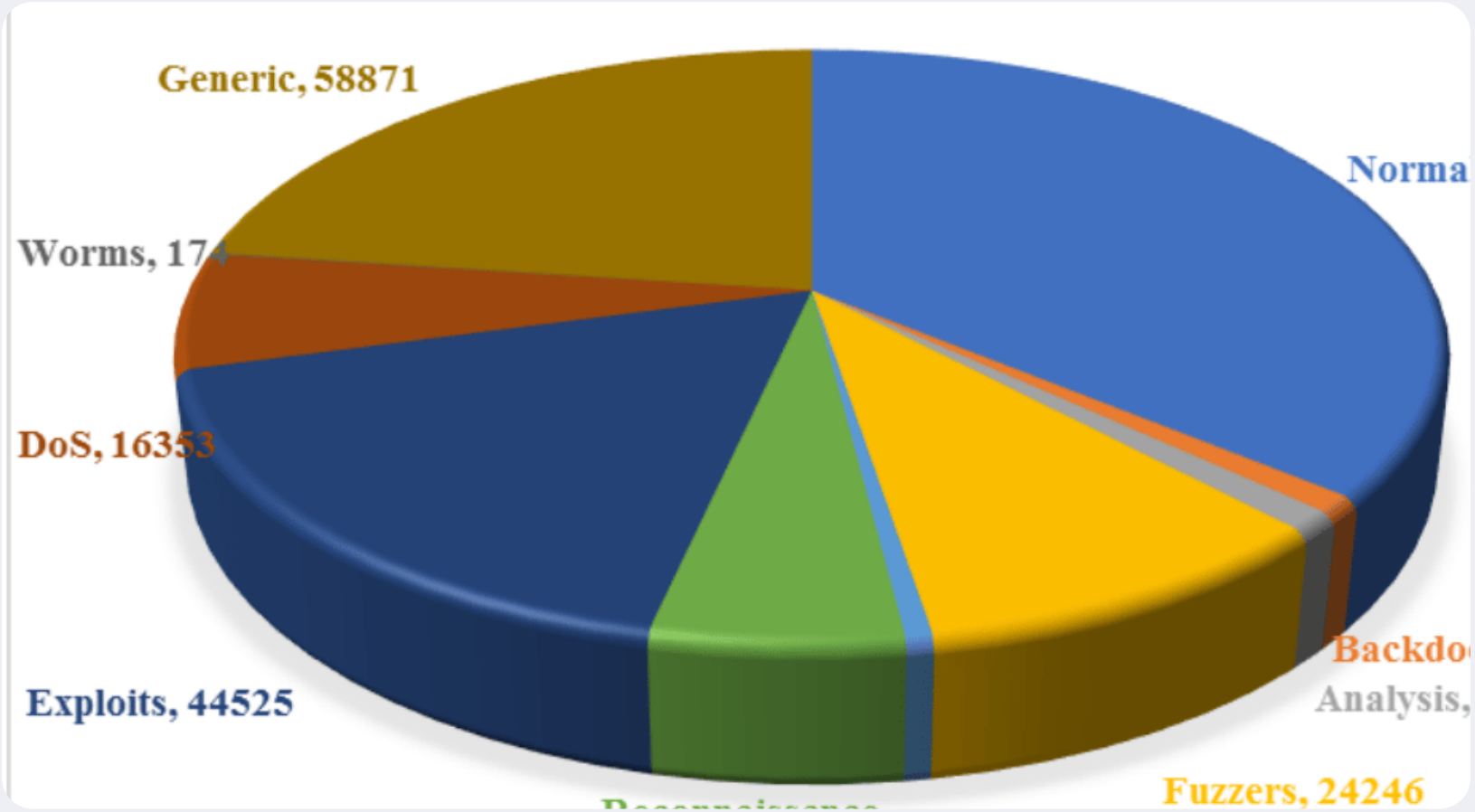
4 Excellent resource for evaluation

With its extensive size and a variety of attacks, UNSW-NB15 is an excellent resource for evaluating and testing the effectiveness of network security algorithms and tools.

5 Ideal for researchers and security professionals

The UNSW-NB15 dataset is an ideal choice for researchers and security professionals who are interested in developing and testing new network security solutions.

Distribution data of UNSW-NB15 data set



Exploratory Data Analysis (EDA)

Before building our intrusion detection system, we need to explore the UNSW-NB15 dataset to gain insights into the data. This process is called Exploratory Data Analysis (EDA).

Visualizing the Data

- One way to gain insights into the data is to visualize it with plots. We used heatmaps to visualize the correlation between features.
- One hot encoding is another technique that we used to encode categorical variables.

Correlation Matrix and PCA

- Another useful tool for EDA is the correlation matrix. It shows the correlation between each pair of features.
- We also performed Principal Component Analysis (PCA) to reduce the dimensionality of the data.

Plan of Work

Plan of Work - Phase 1

1.1 Dataset Selection and Preprocessing:

- Selecting an appropriate dataset to solve the problem statement.
- Handling missing values, balancing the class distribution, and removing noise from the data.

1.2 Feature Selection and Extraction:

- Identifying relevant features that can effectively represent the traffic behavior of IoT devices.
- Extracting features using statistical, frequency-based, or time-domain techniques.

1.3 Algorithm Selection and Implementation:

- Choosing appropriate supervised ML algorithms, like Neural Networks, or other DL algorithms.
- Implementing the selected algorithms using Python libraries, such as Scikit-learn, TensorFlow, etc.

Plan of Work - Phase 2

2.1 Model Training and Evaluation:

- Training the ML models on the preprocessed data using cross-validation techniques.
- Evaluating the performance of the models using various metrics, such as accuracy, precision, recall, and F1-score.

2.2 Deployment and Integration:

- Deploying the trained ML models on IoT devices, gateways, or cloud servers.
- Integrating the IDS with existing security mechanisms, such as firewalls or intrusion prevention systems.

2.3 Optimization and Enhancement:

- Optimizing the ML models for better performance and efficiency.
- Enhancing the IDS by using advanced ML techniques, such as deep learning or ensemble methods.

2.4 Testing and Validation:

- Testing the IDS on real-world IoT networks and assessing its effectiveness in detecting various types of attacks.
- Validating the IDS by comparing its performance with other existing IDSs and analyzing the trade-offs between accuracy and resource consumption.

Model Implementation

In this section, we will discuss the steps and techniques used to implement our machine learning model for intrusion detection in IoT devices. We will cover the following topics:

- The architecture we used for our model **[Neural Network]**
- The preprocessing steps we used to prepare the data
- The evaluation metrics we used to measure the performance of the model

The Architecture of Our Model

For our intrusion detection model, we used a **Neural Network Architecture** that was specifically designed for IoT environments. The architecture consisted of 4 layers of neurons that were connected in a hierarchical manner.

But why Neural Network? Why not something else?

- Neural networks can handle large datasets with thousands or millions of samples.
- Can be easily scaled by increasing the number of layers, neurons, or parameters.
- Can adapt to new or unseen data by generalizing patterns learned from training data.
- Neural networks support end-to-end learning, which means they can learn from raw input data to produce desired output without relying on explicit intermediate steps.

Overall, the architecture of our model was designed to be scalable, efficient, and effective in detecting intrusions in IoT environments.

By using a neural network approach, we were able to achieve high accuracy and reliability, while also being able to adapt to different types of IoT devices and environments.

Preprocessing the Data

In this section, we will discuss the preprocessing steps we used to prepare the data for our intrusion detection model. Preprocessing is a critical step in any machine learning project, and it involves cleaning and transforming the raw data to make it usable for training the model.

First, we loaded the UNSW-NB15 dataset into a Pandas Data Frame and removed any unnecessary columns. We then applied feature scaling to normalize the data and ensure that each feature had equal importance during training.

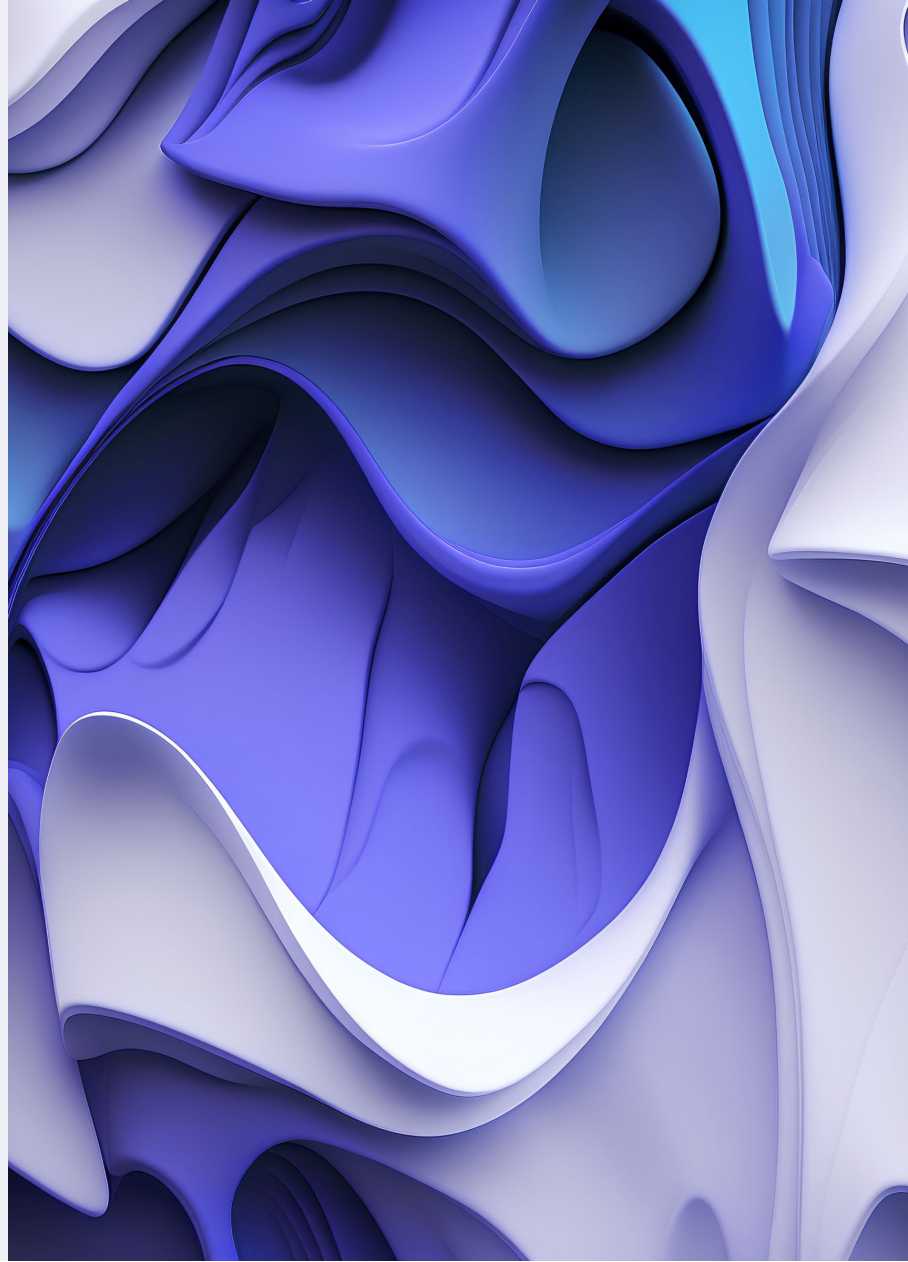
Next, we performed feature engineering to extract relevant features from the raw data. This involved creating new features from existing ones by applying mathematical functions or combining multiple features into a single one.

We also dealt with missing and inconsistent data by imputing missing values and removing any duplicates or outliers. This helped to improve the quality and accuracy of our model.

Overall, our preprocessing steps were designed to ensure that the data was suitable for training our intrusion detection model and that it would produce accurate and reliable results.

Evaluation Metrics

Measuring the effectiveness of our model involves using several evaluation metrics. **Accuracy** measures the percentage of correctly classified samples, while **Train Log loss** measures the error of the model during training. We also use **Value accuracy** to measure the accuracy of the predicted values. Additionally, we analyze the **Confusion matrix**, **Precision matrix**, and **Recall matrix** to evaluate the performance of our model on specific classes. By using these metrics, we can ensure that our intrusion detection system is both accurate and reliable.



Future Scope of Project

In the future, we plan to expand our intrusion detection system to include more IoT devices and types of attacks. We also plan to:

- Investigate more algorithms such as Naïve-Bayes, SVM other DL Algorithms to to further identify which model is better able to identify what types of attacks & why...
- Implement our system on a real-world IoT network to evaluate its performance and scalability.
- Conduct a cost-benefit analysis to determine the feasibility of our system in real-world scenarios.

References

1. C. -Y. Chen, L. -A. Chen, Y. -Z. Cai and M. -H. Tsai, "RNN-based DDoS Detection in IoT Scenario," 2020 International Computer Symposium (ICS), Tainan, Taiwan, 2020.
2. A. Remesh, D. Muralidharan, N. Raj, J. Gopika and P. K. Binu, "Intrusion Detection System for IoT Devices," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020.
3. M. Ge, X. Fu, N. Syed, Z. Baig, G. Teo and A. Robles-Kelly, "Deep Learning-Based Intrusion Detection for IoT Networks," 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC), Kyoto, Japan, 2019.
4. Ravipati, Rama Devi and Abualkibash, Munther, Intrusion Detection System Classification Using Different Machine Learning Algorithms on KDD-99 and NSL-KDD Datasets - A Review Paper (June 2019). International Journal of Computer Science & Information Technology (IJCSIT) Vol 11, No 3, June 2019.
5. Alsulami, Abdulaziz & Abu Al-Haija, Qasem & Tayeb, Ahmad & Alqahtani, Ali. (2022). An Intrusion Detection and Classification System for IoT Traffic with Improved Data Engineering. Applied Sciences.
6. Swarna Sugi, S. S., & Ratna, S. R. (2020). Investigation of Machine Learning Techniques in Intrusion Detection System for IoT Network. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)
7. A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Directions. Javed Asharf , Nour Moustafa , Hasnat Khurshid , Essam Debie , Waqas Haider and Abdul Wahab.