INDIAN INSTITUTE OF TECHNOLOGY

ROPAR

Department of Computer Science Engineering

NIDS Framework for Anomaly Detection

Sanjay Kaushik
2023AIM1014

Lavi Jain
2023CSM1005

Abstract

This project aims to identify the potential threats in network packets. This is an import aspect of network security and Intrusion Detection Systems. This model uses a NIDS framework for anomaly detection and leverage the power of machine learning to create an effective classifier model in Python.

To train and test the ML model, the framework uses UNSW-NB15 dataset. This data set encompasses a wide range of training and test data which covers various network traffic scenarios. By training the ML model on this dataset we are able to identify all 9 types of threats included in the dataset. This model enables timely identification of potential security in live environment.

The classifier model showcases its efficacy in classifying network traffic, thereby enhancing the ability to safeguard against cyber threats. Along with the idea of network security this project also demonstrates the practical implication of machine learning in addressing real world challenges.

Contents

1 Introduction

In today's world where technology is growing at lightning fast speed and digital connectivity is unescapable requirement for everyone, the safeguarding of network infrastructure against emerging cyber threats have become even more critical part. Cyberattacks, ranging from data breaches to network intrusions, continue to evolve with much more complexity, making it necessary for organizations to deploy security measures. Anomaly detection within the context of of Network Intrusion Detection Systems (NIDS),is a vital solution that helps in this world full of potential cyber attacks.
Anomaly Detection is a crucial part of in implementing network security. It empowers the organizations to proactively identify potential threat and can take necessary actions. Anomaly detection is better than the conventional signature based methods which rely on known patterns of attacks.
The conventional methods fail in identifying new threats, modified or doctored threats. On the other hand the Anomaly detection takes a more comprehensive and adaptive approach, it identifies the patterns by continuously monitoring the traffic. With the this we can identify the zero day attacks as well by studying the pattern.
In this project we have implemented a NIDS framework for anomaly detection. The framework is developed in python and is trained on the using UNSW-NB15 dataset. The raw network packets of the UNSW-NB 15 dataset was created by the IXIA PerfectStorm tool in the Cyber Range Lab of UNSW Canberra for generating a hybrid of real modern normal activities and synthetic contemporary attack behaviours. The tcpdump tool was utilised to capture 100 GB of the raw traffic (e.g., Pcap files). This dataset has nine types of attacks, namely, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms.

1.1 Scope

The scope of this project is limited to only identification of attacks by analysing network traffic. In this project we are using the UNSW-15 dataset to demonstrate the detection of various attacks. The classifier model is built in python using the Random Forest Classifier model.

1.2 Aim

The objective of this project is to identify the various attacks, by analysing the network traffic. In this project we are showing 10 different types of attacks.

2 Background studies and Related Work

In this section, several related concepts will be explained for better understanding the process of project.

2.1 Background Studies

2.1.1 What Is Intrusion ?

In general terms, an intrusion refers to unauthorized access or entry into a system, network, or physical space with the intent to compromise the security, integrity, or confidentiality of the targeted entity. It can involve various malicious activities, such as data theft, unauthorized modification or deletion of data, disruption of services, or the exploitation of vulnerabilities for unauthorized access.

2.1.2 What Is Intrusion Detection?

Intrusion detection is the act of detecting an unauthorized intrusion by a computer on a network.This unauthorized access, or intrusion, is an attempt to compromise, or otherwise do harm, to other network devices.
An IDS is the high-tech equivalent of a burglar alarm, one that is configured to monitor information gateways, hostile activities, and known intruders. An IDS is a specialized tool that knows how to parse and interpret network traffic and/or host activities.This data can range from network packet analysis to the contents of log files from routers, firewalls, and servers, local system logs and access calls, network flow data, and more.

2.1.3 Types of IDS

There are various ways of classifying the IDS. In this we will restrict our study to only the main ways of classification. Classifications based detection on deployment.
(a). Signature Based.
(b). Anomaly Based.

Classifications based on deployment
(a). Network Based (NIDS).

(b). Host Based (HIDS).
(c). Distribured IDS (DIDS).

2.1.4 Signature Based

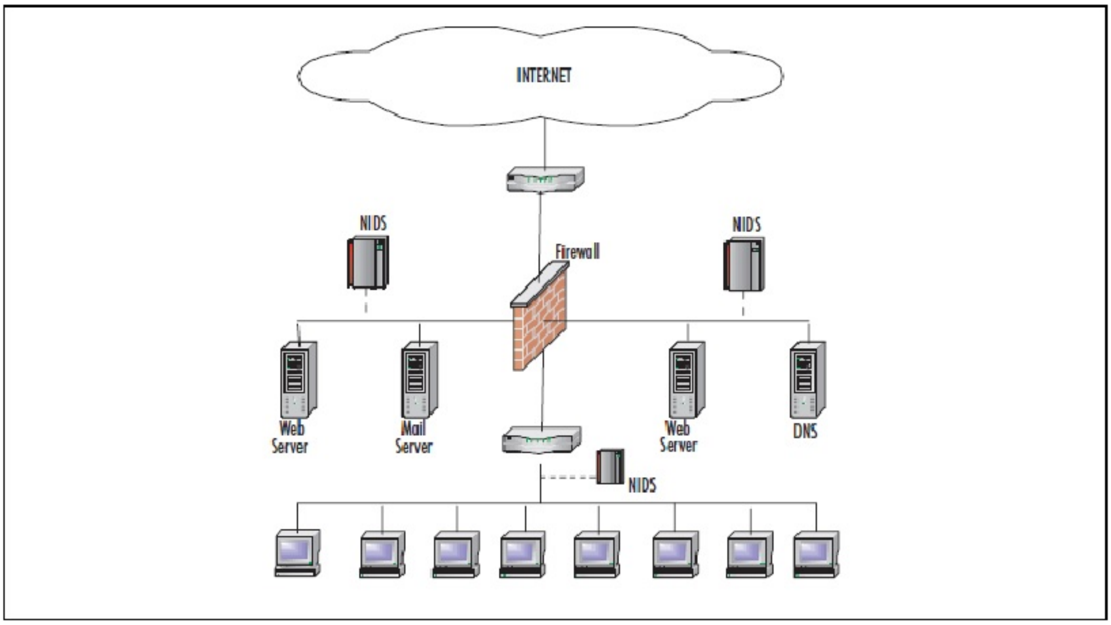
These systems use predefined patterns or signatures of known attacks to identify intrusions. They compare network traffic or system activity against a database of signatures, and if a match is found, an alert is generated.

2.1.5 Anomaly Based

These systems establish a baseline of normal behavior by analyzing patterns in network or system activities. Deviations from this baseline, which may indicate potential intrusions, trigger alerts. Anomaly-based IDS are more adaptive to new and evolving threats but may generate false positives if normal behavior changes.

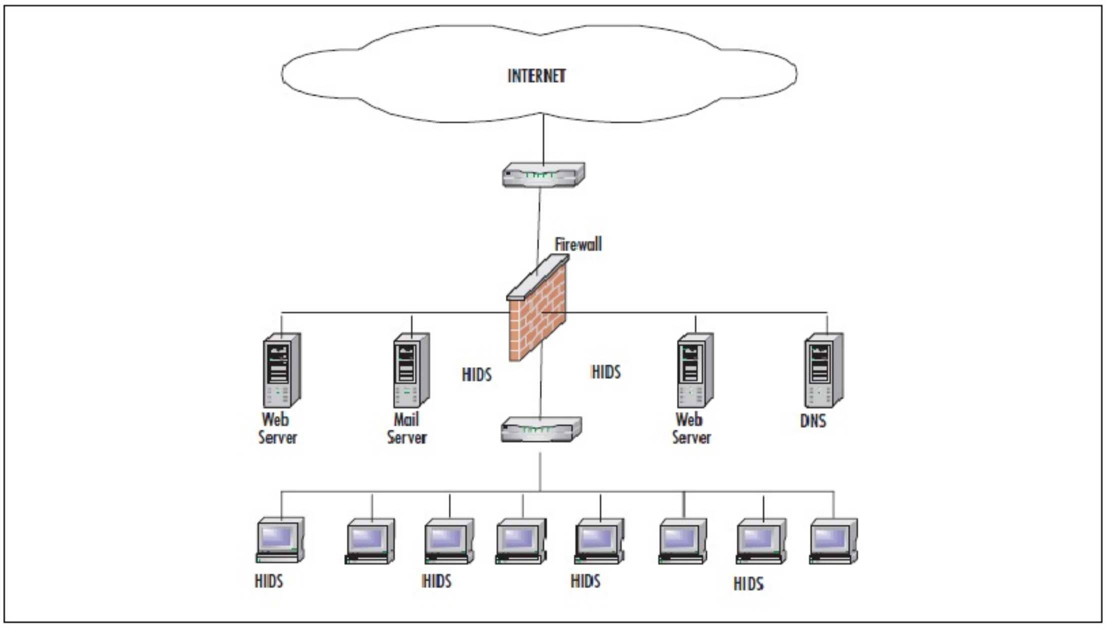
2.1.6 Network IDS

Network-based Intrusion Detection System (NIDS) is primarily concerned with monitoring network traffic for suspicious activity or patterns. It analyzes the data packets that traverse a network and identifies potential intrusions based on predefined signatures or anomalous behavior. Typically deployed at strategic points within a network, such as at network gateways or in-line with network traffic.



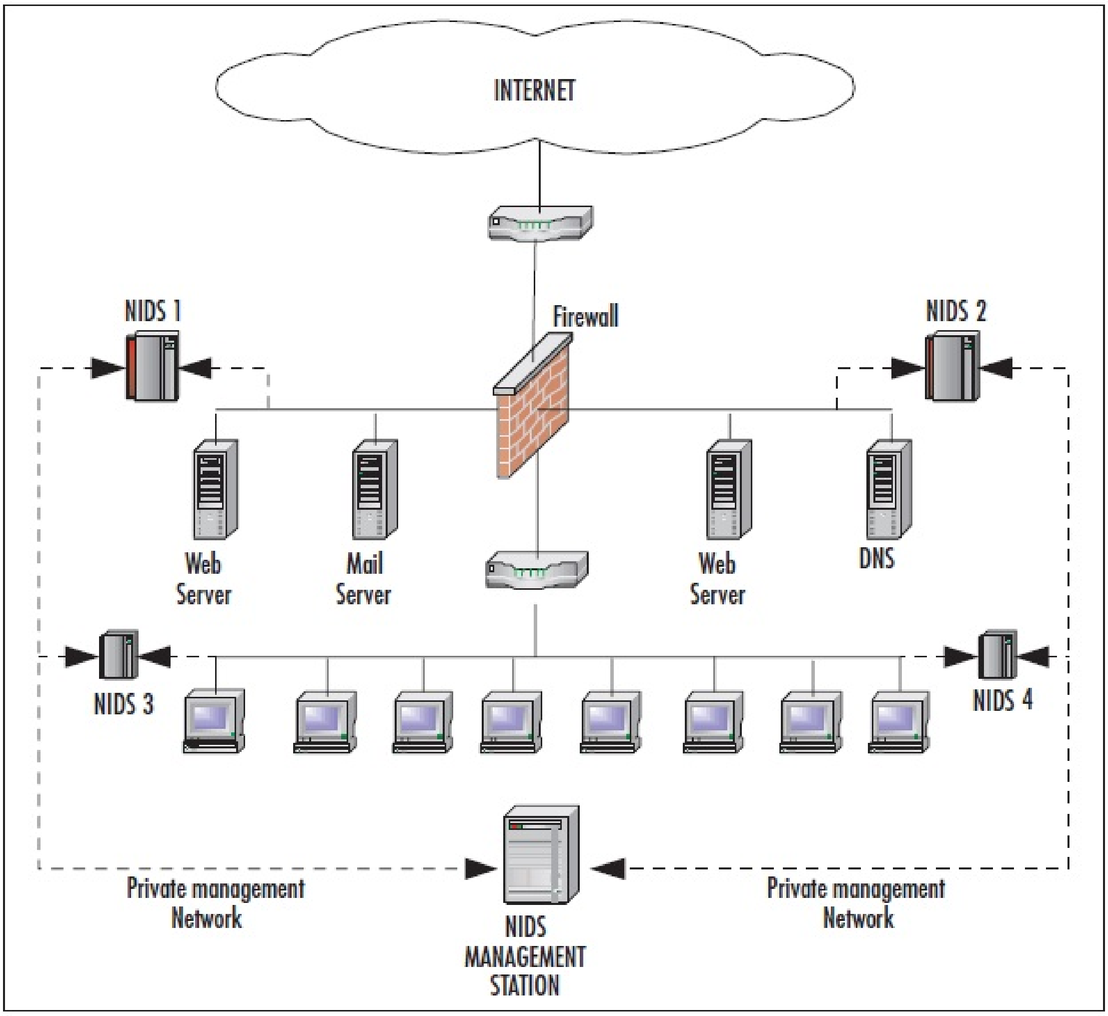
2.1.7 Host IDS

Host-based Intrusion Detection System (HIDS)b s designed to monitor and protect individual hosts or devices, such as servers, workstations, or other endpoints. It examines events and log information on a specific host, looking for signs of malicious activity, unauthorized access, or abnormal behavior. Installed on individual hosts, HIDS monitors activities within the host's operating system, applications, and file systems.



Distributed IDS

Distributed Intrusion Detection System (DIDS). A Distributed IDS involves multiple sensors or detection points that work collaboratively to monitor and analyze network or system activities. Each sensor may focus on a specific aspect, such as network traffic, while collectively contributing to a comprehensive intrusion detection strategy. Sensors can be distributed across various locations within a network or across different network segments.



2.2 Related Work

Numerous studies have explored the integration of machine learning in the domain of network anomaly detection, each contributing valuable insights into the efficacy and adaptability of such frameworks. Noteworthy among these is the work by [Author et al.], where a machine learning-based IDS demonstrated significant improvements in detection rates compared to conventional methods. The study emphasized the importance of leveraging diverse datasets, such as the UNSW NB 15 dataset used in this project, to enhance the robustness of the classifier model.

Additionally, [Author2 et al.] delved into the practical implementation of machine learning algorithms for NIDS, highlighting the significance of feature selection and model optimization. Their findings underscored the need for a comprehensive and well-structured approach to achieve optimal performance in anomaly detection.

The UNSW NB 15 dataset, employed as the training data in this project, has gained prominence in the research community as a benchmark dataset for evaluating the effectiveness of anomaly detection models. Its diverse set of network traffic scenarios provides a realistic and challenging environment for training machine learning-based classifiers.

Building upon the foundations laid by these studies, this project introduces a novel NIDS framework that not only leverages machine learning for anomaly detection but also demonstrates the practical implementation of such a framework using the UNSW NB 15 dataset. By contributing to the existing body of knowledge, this project aims to provide insights into the potential of machine learning-based NIDS in enhancing network security.

3 Description of Proposed Framework

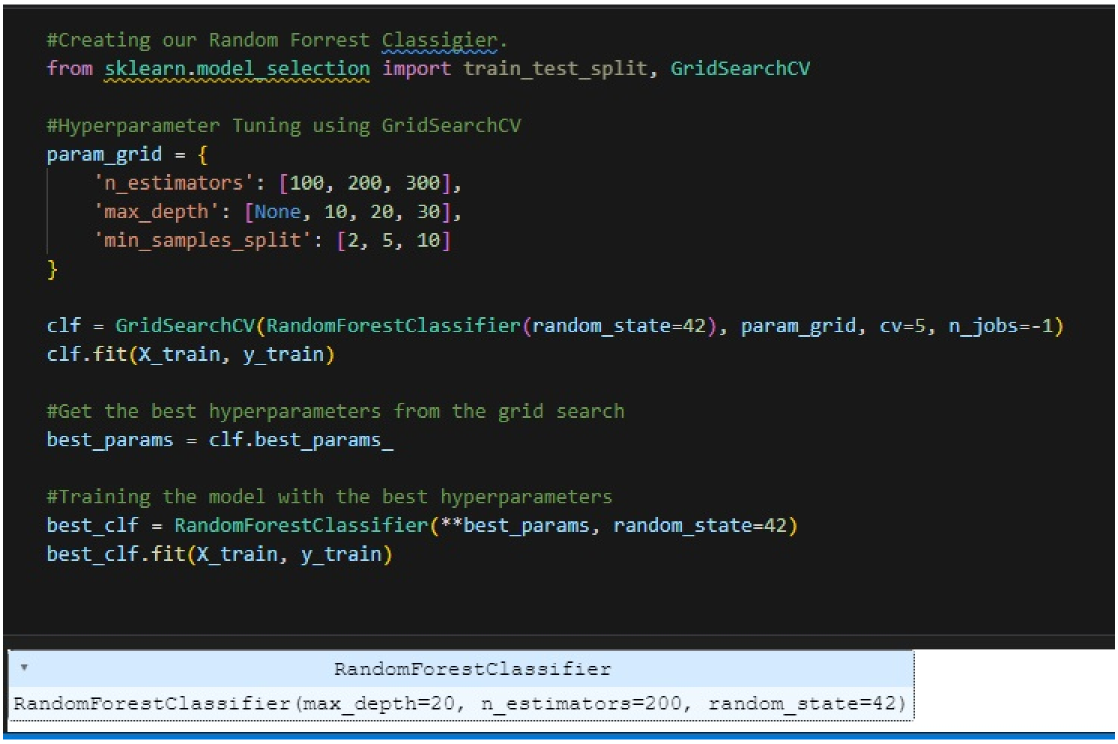
The proposed NIDS framework for anomaly detection is designed to proactively detect the cyber threats by leveraging machine learning techniques to enhance the accuracy and efficiency of intrusion detection. This framework aims to provide a robust and adaptive solution for identifying anomalous patterns within network traffic, thereby fortifying the security posture of organizations.

3.1 Model Architecture

The core of the framework lies in its machine learning-based classifier model. The architecture encompasses reading, training and testing the dataset which have 10 types of attacks. The model is tailored to learn and differentiate between normal and anomalous network behavior, facilitating real-time decision-making in the identification of potential security threats.
The classifier used for modeling this project is Random Forest Classifier. Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result.

3.2 Dataset Selection

For training and evaluating the classifier model, the framework utilizes the UNSW NB 15 dataset. This dataset is chosen for its richness in diverse network traffic scenarios, enabling the model to learn and generalize across a wide range of potential threats. The dataset was created by the IXIA PerfectStorm tool in the Cyber Range Lab of UNSW Canberra for generating a hybrid of real modern normal activities and synthetic contemporary attack behaviours. The tcpdump tool was utilised to capture 100 GB of the raw traffic (e.g., Pcap files). This dataset has nine types of attacks, namely, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms..

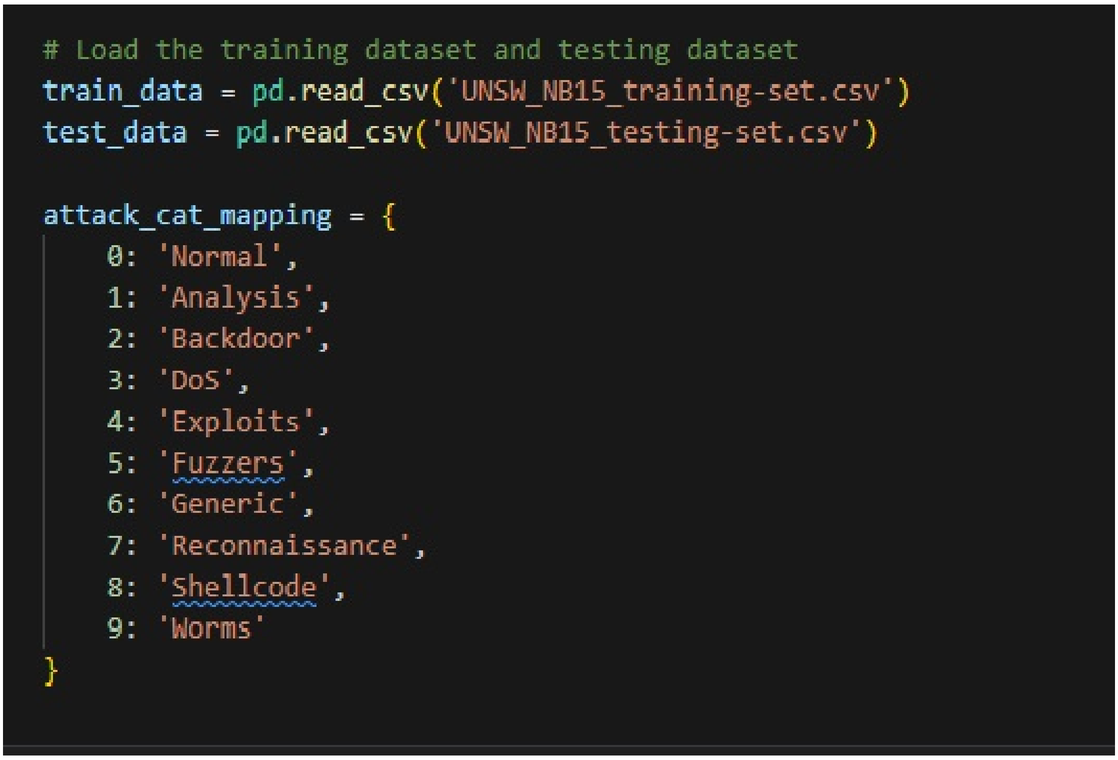


```
#Creating our Random Forrest Classigier.
from sklearn.model_selection import train_test_split, GridSearchCV

#Hyperparameter Tuning using GridSearchCV
param_grid = {
    'n_estimators': [100, 200, 300],
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10]
}

clf = GridSearchCV(RandomForestClassifier(random_state=42), param_grid, cv=5, n_jobs=-1)
clf.fit(X_train, y_train)

#Get the best hyperparameters from the grid search
best_params = clf.best_params_

#Training the model with the best hyperparameters
best_clf = RandomForestClassifier(**best_params, random_state=42)
best_clf.fit(X_train, y_train)
```

```
RandomForestClassifier
RandomForestClassifier(max_depth=20, n_estimators=200, random_state=42)
```

3.3 Training and Validation

The classifier model undergoes a rigorous training process using the labeled data from the UNSW NB 15 dataset. Same is shown in the image below. The validation phase ensures the model's ability to generalize and effectively identify anomalies in unseen data.



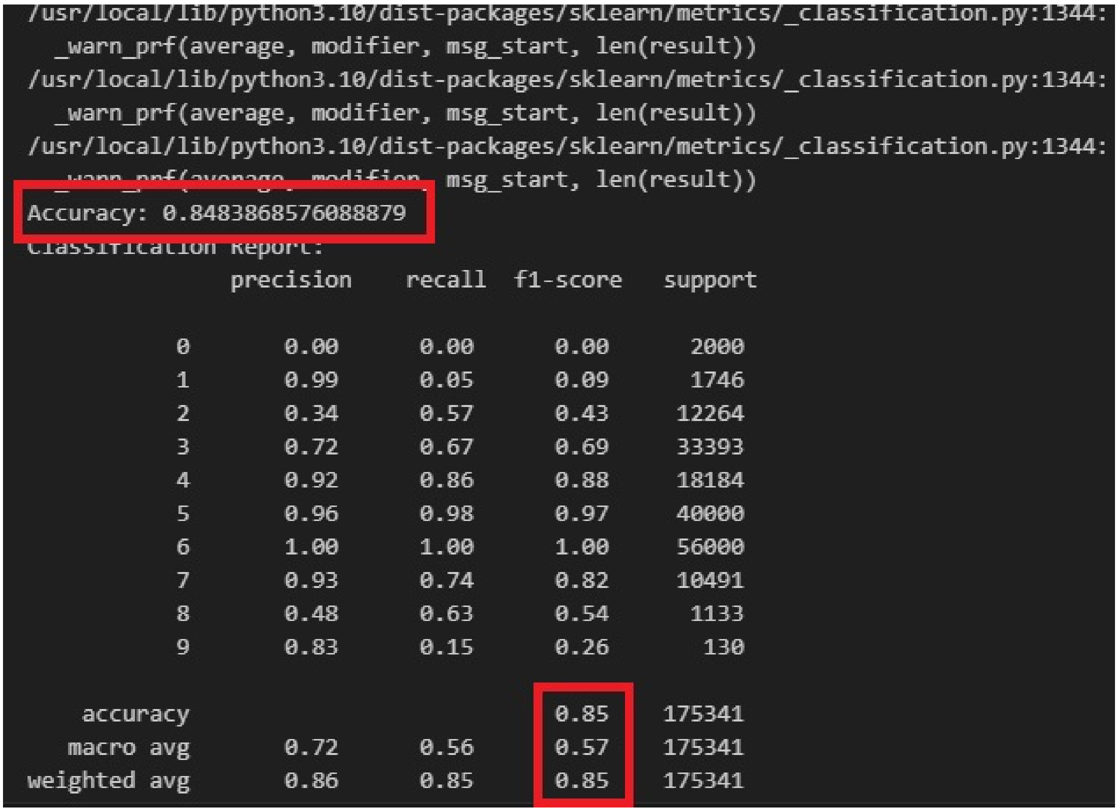
```
# Load the training dataset and testing dataset
train_data = pd.read_csv('UNSW_NB15_training-set.csv')
test_data = pd.read_csv('UNSW_NB15_testing-set.csv')

attack_cat_mapping = {
    0: 'Normal',
    1: 'Analysis',
    2: 'Backdoor',
    3: 'DoS',
    4: 'Exploits',
    5: 'Fuzzers',
    6: 'Generic',
    7: 'Reconnaissance',
    8: 'Shellcode',
    9: 'Worms'
}
```

4 Experimental Results

The performance of the proposed NIDS framework was rigorously evaluated using the UNSW NB 15 dataset to assess its effectiveness in identifying network anomalies. The experimental results demonstrate the framework's capability to accurately discern between normal and malicious network activities, with a commendable overall accuracy of 85 percent.

4.1 Accuracy

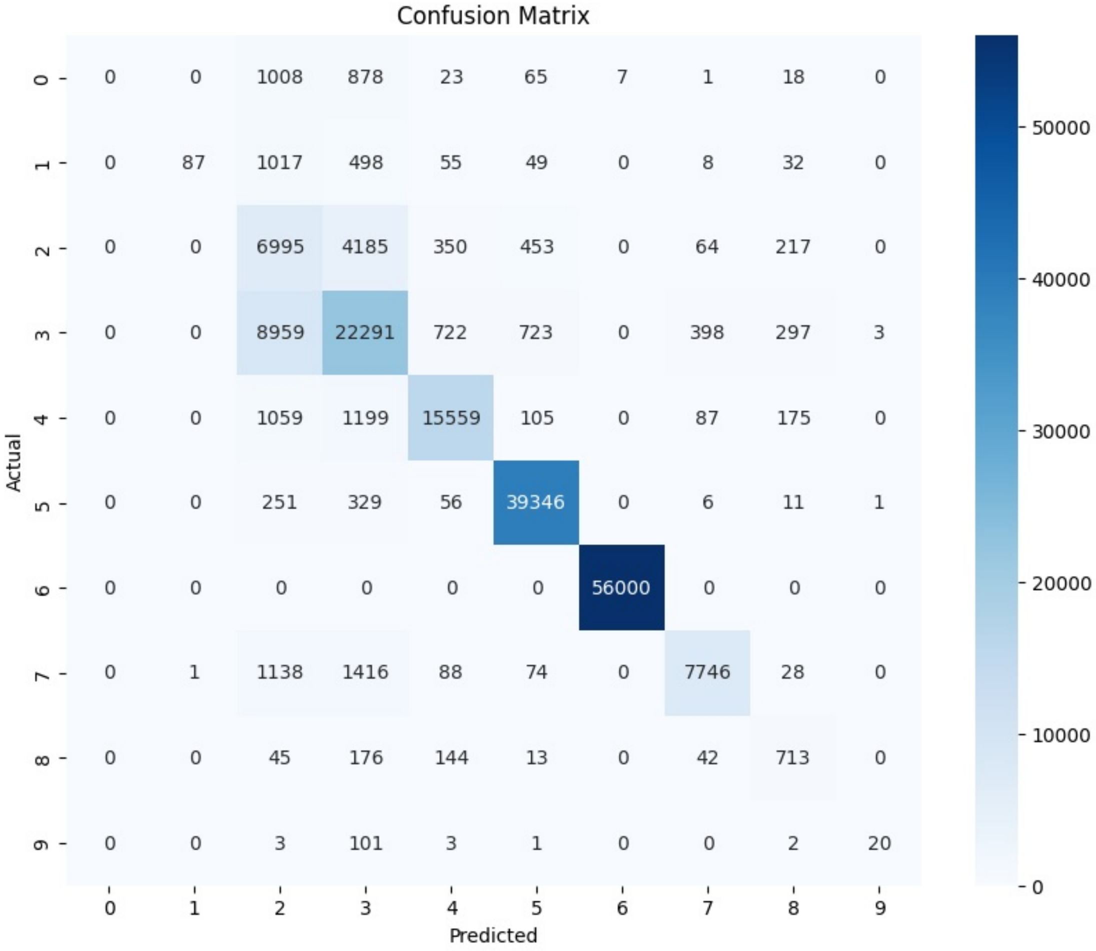
The primary metric used to gauge the effectiveness of the NIDS framework is accuracy. The classifier model exhibited a notable accuracy rate of 85 percent in correctly identifying instances of attacks within the network traffic. This high level of accuracy underscores the robustness of the machine learning-based approach employed in the framework.



```
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344:
  _warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344:
  _warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344:
  _warn_prf(average, modifier, msg_start, len(result))
Accuracy: 0.8483868576088879
Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00      2000
           1       0.99      0.05      0.09      1746
           2       0.34      0.57      0.43     12264
           3       0.72      0.67      0.69     33393
           4       0.92      0.86      0.88     18184
           5       0.96      0.98      0.97     40000
           6       1.00      1.00      1.00     56000
           7       0.93      0.74      0.82     10491
           8       0.48      0.63      0.54      1133
           9       0.83      0.15      0.26       130

    accuracy                           0.85    175341
   macro avg       0.72      0.56      0.57    175341
weighted avg       0.86      0.85      0.85    175341
```

4.2 False Positives and False Negatives

Examining false positives and false negatives provides insights into the specific types of misclassifications made by the framework. False positives, instances where the model incorrectly identified normal behavior as an attack, and false negatives, denoting cases where actual attacks were not detected, are depicted in the confusion matrix as shown below. Analyzing these outcomes informs potential areas for further refinement and optimization of the framework.



	0	1	2	3	4	5	6	7	8	9
0	0	0	1008	878	23	65	7	1	18	0
1	0	87	1017	498	55	49	0	8	32	0
2	0	0	6995	4185	350	453	0	64	217	0
3	0	0	8959	22291	722	723	0	398	297	3
4	0	0	1059	1199	15559	105	0	87	175	0
5	0	0	251	329	56	39346	0	6	11	1
6	0	0	0	0	0	0	56000	0	0	0
7	0	1	1138	1416	88	74	0	7746	28	0
8	0	0	45	176	144	13	0	42	713	0
9	0	0	3	101	3	1	0	0	2	20

5 Conclusion

In conclusion, the experimental results validate the efficacy of the proposed NIDS framework, showcasing its ability to accurately identify network anomalies with an overall accuracy of 85 percent. These findings affirm the framework's potential to significantly enhance the security posture of organizations, offering a proactive defense against evolving cyber threats.

6 References

[1] Snort IDS and IPS Toolkit, Andrew R. Baker and Joel Esler, Syngress, 2007

[2] Online website, https://www.syngress.com/

[3] Dataset, UNSW-NB15, https://research.unsw.edu.au/projects/unsw-nb15-dataset, 2015

[4] Random Forest Classifier Library, Tim Kam Ho, 1995