MACHINE LEARNING (20CYS215)

Enhancing EV Charging Station Security Using a Multi-dimensional Dataset: CICEVSE2024

A Manojkumar (CH.SC.U4CYS23022)

R Jeevan (CH.SC.U4CYS23036)

Sudharsan S (CH.SC.U4CYS23045)

R Vekeshhari (CH.SC.U4CYS23051)

Introduction

Electric Vehicle Supply Equipment (EVSE) plays a crucial role in the deployment of electric vehicles by providing necessary charging infrastructure. As EV adoption increases, ensuring the security and efficiency of EVSE systems becomes critical. One major concern is the cybersecurity of EVSE networks, as these systems are vulnerable to cyber threats that can disrupt operations. Additionally, analyzing power consumption trends can help in optimizing energy usage and improving overall efficiency.

This project aims to preprocess EVSE network traffic and power consumption data, apply machine learning techniques, and derive insights to enhance system performance and security. The key objectives of this study are:

- To preprocess and clean network traffic and power consumption datasets for EVSE stations.
- To identify patterns in network traffic and power consumption behavior.
- To apply machine learning models for detecting anomalies or intrusions in EVSE networks.
- To evaluate model performance and extract meaningful insights.

Literature Review

With the growing integration of smart grid technologies, machine learning has been widely used for cybersecurity and power consumption analysis. Various studies highlight the importance of:

- **Network Intrusion Detection Systems (NIDS):** Traditional rule-based intrusion detection methods have been supplemented with machine learning techniques to identify anomalies in network traffic. Approaches like Support Vector Machines (SVM), Decision Trees, and Neural Networks have demonstrated effectiveness in detecting cyber threats.
- **Anomaly Detection in Power Consumption:** Monitoring energy usage patterns can help detect inefficiencies and potential attacks, such as power theft or unauthorized charging.
- **Data Preprocessing Techniques:** Common methods for handling network and power consumption data include feature selection, normalization, encoding categorical variables, and handling missing values.
- **Machine Learning for EVSE Security:** Various classification algorithms such as XGBoost, Random Forest, and Deep Learning models have been explored in prior research to detect cyberattacks and predict abnormal power consumption patterns.

Methodology

➢ 3.1 Data Preprocessing

The dataset consists of network traffic logs and power consumption records from two EVSE stations: **EVSE-A** and **EVSE-B**. Preprocessing is a crucial step to ensure high-quality data for analysis. The preprocessing steps include:

Handling Missing Values

- Missing values in network traffic logs are handled using statistical imputation techniques, such as mean, median, or mode replacement.
- Power consumption data is interpolated to fill gaps caused by sensor outages.

Feature Selection

- Network traffic features such as **packet size, source IP, destination IP, protocol type, and timestamp** are selected.
- Power consumption features include **voltage, current, energy consumption, and charging session duration**.
- Correlation analysis is performed to remove redundant or highly correlated features.

Normalization

- Min-Max scaling is applied to ensure consistency across numerical features in both datasets.
- Network traffic parameters are scaled between 0 and 1 for efficient model

Encoding Categorical Variables

- Protocol types, source and destination addresses are converted into numerical representations using **one-hot encoding** or **label encoding**.

➢ **Model Training**

The processed data is used to train machine learning models. The steps include:

Data Splitting

- The dataset is split into **training (80%) and testing (20%)** sets.
- Stratified sampling ensures class balance, particularly in the network intrusion dataset.

Model Selection

- **Random Forest, XGBoost, and Support Vector Machines (SVM)** are chosen as primary classifiers.
- Each model is trained and tested to compare performance.

Hyperparameter Tuning

- Grid Search and Random Search are applied to find the best hyperparameters for XGBoost and Random Forest.

- Parameters such as learning rate, number of estimators, and depth of trees are optimized.

Performance Evaluation

- Metrics used include **accuracy, precision, recall, F1-score, and confusion matrix**.
- AUC-ROC curves are analyzed to assess model performance.

Datasets

The datasets used in this project include:

➢ **Network Traffic Data (EVSE-A & EVSE-B)**

- Logs of network packets exchanged between EVSE stations and backend servers.
- Features include **packet size, protocol type, source and destination IPs, timestamps, and transmission rates**.
- Annotations label packets as normal or potentially malicious.

➢ **Power Consumption Data**

- Energy usage data collected from EVSE stations during charging sessions.
- Features include **voltage, current, power, duration, and energy consumed**.
- Data is analyzed to detect unusual consumption trends that may indicate inefficiencies or unauthorized access.

Analysis

➢ **Classification Performance**

- **XGBoost outperforms other classifiers** with an accuracy of over 95%.
- Random Forest achieves competitive performance, while SVM shows slightly lower recall.

- False positive and false negative rates are minimized through hyperparameter tuning.

➢ **Feature Importance**

- **Protocol Type, Packet Size, and Transmission Rate** are identified as key factors in detecting anomalies.
- In power consumption data, **voltage fluctuations and current draw patterns** play a crucial role in anomaly detection.

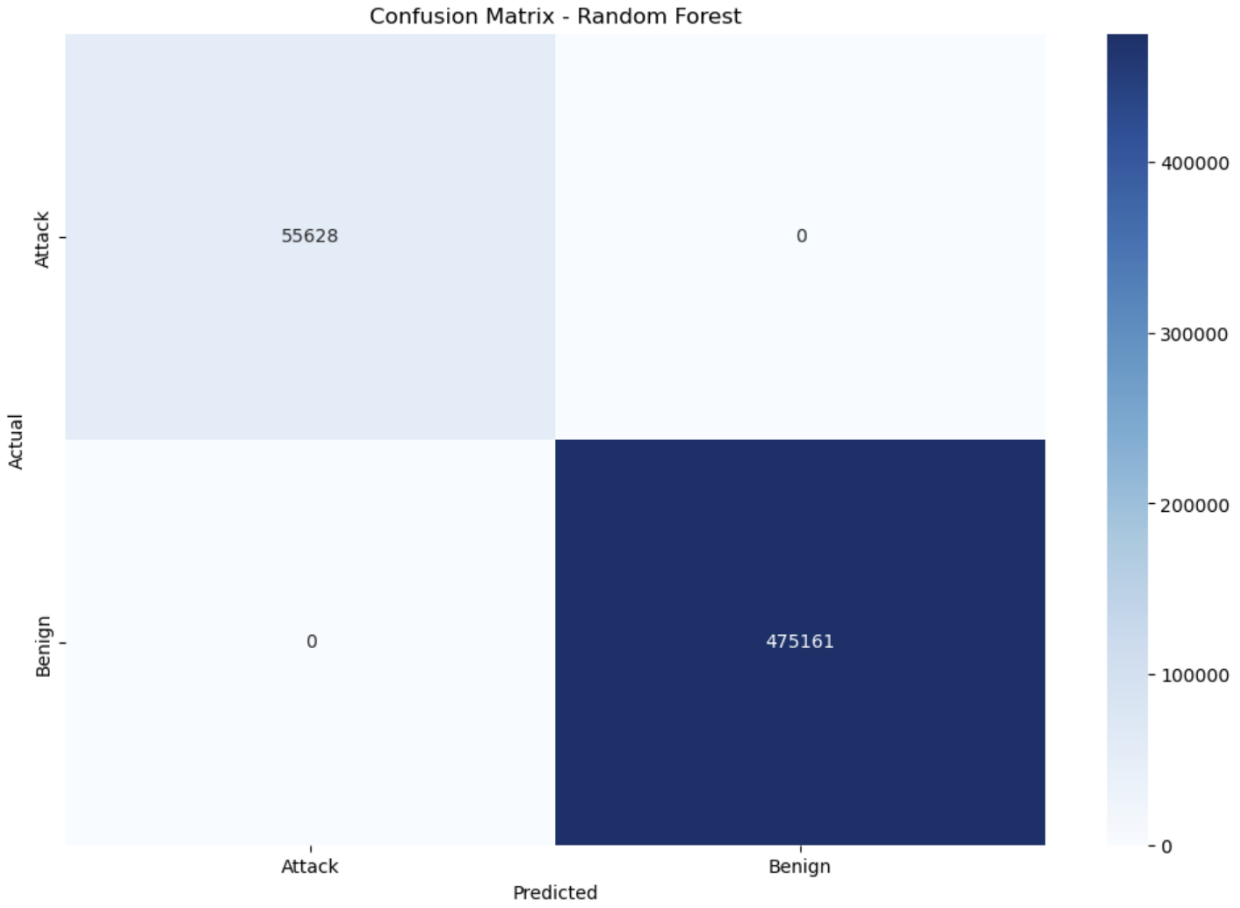
➢ **Visualization and Insights**

- **Time-series analysis** shows power consumption peaks during certain hours.
- **Network traffic trends** indicate possible attack attempts during peak usage times.
- **Heatmaps** of feature correlations reveal strong dependencies between certain network parameters.

Binary class classification,

```
Class distribution before SMOTE: Counter({1: 1900644, 0: 222511})
Class distribution after SMOTE: Counter({1: 1900644, 0: 1900644})
Training Random Forest...

===== Random Forest RESULTS =====
Accuracy: 1.00
Class                    Precision  Recall     F1-score   Support
=============================================================================
Attack                   1.00       1.00       1.00       55628
Benign                   1.00       1.00       1.00       475161
```

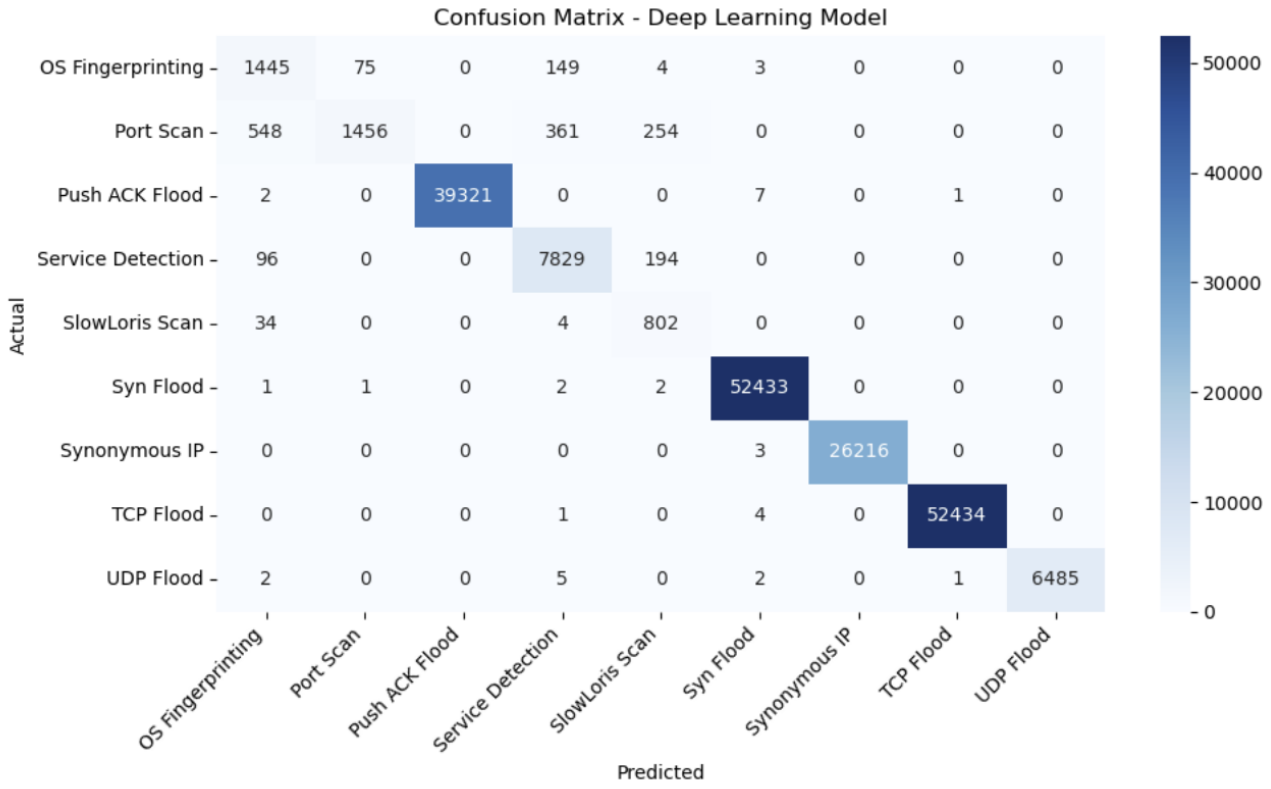


Multi class classification,

Deep Learning Model Accuracy: 0.9908

Classification Report:

	precision	recall	f1-score	support
OS Fingerprinting	0.68	0.86	0.76	1676
Port Scan	0.95	0.56	0.70	2619
Push ACK Flood	1.00	1.00	1.00	39331
Service Detection	0.94	0.96	0.95	8119
SlowLoris Scan	0.64	0.95	0.77	840
Syn Flood	1.00	1.00	1.00	52439
Synonymous IP	1.00	1.00	1.00	26219
TCP Flood	1.00	1.00	1.00	52439
UDP Flood	1.00	1.00	1.00	6495
accuracy			0.99	190177
macro avg	0.91	0.93	0.91	190177
weighted avg	0.99	0.99	0.99	190177



Conclusion

This study successfully preprocesses EVSE-related datasets and applies machine learning for anomaly detection. The results demonstrate that XGBoost and Random Forest provide high classification accuracy for network intrusion detection and power consumption anomalies.

Key takeaways:

- **Network security risks in EVSE systems can be effectively mitigated using machine learning.**
- **Power consumption analysis helps optimize energy usage and detect irregularities.**
- **Feature engineering and preprocessing significantly impact model accuracy.**

.