An Intelligent Data-Driven Model to Secure Intravehicular Communications Based on Machine Learning

Sairam Kodimella
Mentors: Dr. Christelle Scharff
Seidenberg Graduate Research
Pace University, Seidenberg School of CSIS

GitHub – https://bit.ly/3AlAqXX
Notebook (Model) – https://bit.ly/3oVn6Nh

Abstract

Electric vehicular usage gaining more popularity because of intra-vehicular communication between various sensors and hackers find more ways to intercept and modify the functionality of the algorithms used for reliable intra-vehicular communication between sensors. Through this project we propose a secure and reliable intelligent framework to prevent hackers from attacking the vehicles. The proposed method uses an enhanced support vector machine model for anomaly detection, based on the controller area network (CAN) bus protocol. The simulation results using actual data sets demonstrate the effectiveness, dependability, and security of the proposed model in protecting against denial-of-service (DoS) attacks on electric vehicles.

Introduction

There is a piece of hardware called electronic control units (ECUs) in vehicles and they will be communicating through network protocols such as CAN, LIN, FlexRay or MOST. However, the CAN bus protocol which is commonly used in vehicles suffers from security issues in the new dynamic environment of smart grids, making electric vehicles vulnerable to cyber attacks. Many researchers have proposed many methods for detecting and preventing such attacks, including intrusion detection systems, data management systems, and firewalls. The method discussed in the project proposes an intelligent and highly secure method based on support vector machines and one-class detection systems to equip electric vehicles with a powerful anomaly detection and avoidance mechanism. The proposed method uses experimental CAN bus data and an optimization algorithm based on social spider optimization to adjust the SVR setting parameters. A two-stage modification method based on genetic algorithms is developed to increase population diversity and avoid premature convergence. The proposed model's feasibility and performance are examined using real datasets gathered from an electric vehicle.

Research Questions

- In comparison to traditional methods such as AES encryption and the CONTROLLER AREA NETWORK (CAN) protocol, how effective is the implementation of machine learning algorithms in preventing malicious intruder attacks on the communication between sensors and Electronic Control Units (ECU) in electric vehicles?

Related Work

- Zhang et al. (2017), "Security Issues and Challenges in Connected Electric Vehicles" - This article examines the security difficulties and challenges in connected electric cars, as well as a security architecture for ensuring data privacy and integrity.
- Tian et al. (2018) - "Intrusion detection system for automotive networks: a machine learning approach" This study describes an intrusion detection system for automotive networks that is based on machine learning methods like decision trees and SVM.

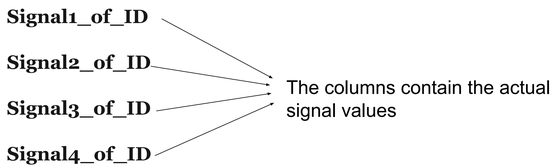
Dataset

The purpose of this dataset, which we name SynCAN Dataset (Synthetic CAN Bus Data), is to offer a baseline for evaluating and comparing different CAN Intrusion Detection Systems (IDS) on diverse signal space assault scenarios. The data set consists of a training data set and test data sets that all contain the following columns:

Label - The column 'Label' indicates whether the data row is to be considered as normal (Label=0) or as intrusion (Label=1).

ID - The column 'ID' contains the identifiers for the IDs that are 'id1', ..., 'id10'.

Time - In the column 'Time' the time stamp of the current message is represented in milliseconds.



Methodology

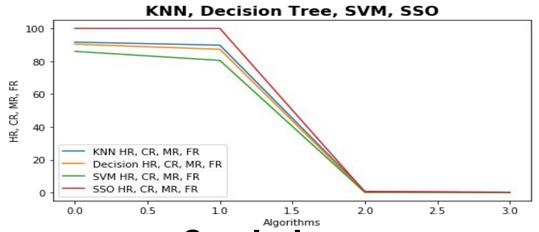
- **Loading Dataset**: The code loads the dataset from the given file path and displays the values.
- **Dataset Exploration**: The code explores the dataset by describing each column in terms of count, mean, standard deviation, etc., and finding the count of different signals on different IDs.
- **Dataset Visualization**: The code performs data visualization by finding and plotting graphs of attacks from the dataset and signal graphs with and without attacks.
- **Dataset Preprocessing**: The code preprocesses the dataset by converting non-numeric data into numeric values using Label Encoder and replacing missing values with mean.
- **Model Building**: The code then builds three different models for classification, including SVM, KNN, and Decision Tree, using genetic selection for feature selection. It splits the dataset into training and testing sets and trains the models on the training set. The accuracy, precision, and confusion matrix are then calculated on the testing set.
- **Model Evaluation**: Finally, the code evaluates the models based on their accuracy, precision, and confusion matrix and selects the best-performing model.

Experiments

The results yielded after uploading the dataset as is differ much significantly after we experimented for examining to check if any patterns or anomalies by EDA, we have then removed all the null values and made non-numeric to numeric then the hit-rate or the accuracy of the model increased significantly

Result

After performing all the above steps listed in methodologies, we have supplied the train and test data to different algorithms which yielded different results, among them SSO algorithm after using features selected by Genetic algorithm has higher hit rate.



Conclusion

This proposed project a novel intelligent and secured anomaly detection model for cyberattack detection and avoidance in the electric vehicles. From the cyber security point of view, the proposed model could successfully detect malicious behaviors while letting the trusted message frames broadcast in the CAN protocol. The high HR% and FR% indices prove the true positive and true negative decisions made by the proposed model. Regarding the MR% and CR% indices, the very low values which most of them are around the upper and lower bounds of the message frame frequency, show the highly trustable performance of this model. The project will assess the effect of other cyberattacks on the performance of different anomaly detection models in the future works.

Future Work

Developing a real-time intrusion detection system: The current project focuses on offline detection of intrusions using a pre-collected dataset. However, a real-time intrusion detection system that can detect intrusions as they happen could be more useful in preventing unauthorized access.

Improving the accuracy of the model: The accuracy of the intrusion detection model could be improved by using more advanced machine learning techniques or incorporating other data sources.

Testing the model on different types of vehicles: The current project only uses data from a specific type of vehicle. Testing the model on different types of vehicles could determine if the model's accuracy varies depending on the type of vehicle.