Machine Learning-Driven Distributed Intrusion Detection System for the Internet of Vehicles

Joud AlFarra[1], Leena Alam[1], Naila Marir[1,2] and Akila Sarirete[1]

[1]Computer Science Department, Effat College of Engineering, Effat University, Jeddah, Saudi Arabia
[2]LIRE Laboratory, Constantine 2-Abdelhamid Mehri University, Constantine, Algeria

Abstract

The Internet of Vehicles (IoV), integrating vehicles with EV charging stations, faces significant vulnerabilities to cyber threats. Designing an efficient Intrusion Detection System (IDS) to detect sophisticated attacks within EV charging stations is crucial for safeguarding this interconnected network. However, creating IDS models using centralized machine learning approaches requires overcoming challenges such as scalability, time-consuming analysis, and real-time processing requirements. In this context, we propose a novel distributed system to IDS design for the IoV, leveraging distributed learning techniques to address these challenges and enhance EV charging station security. This study utilizes different machine learning models to improve the effectiveness of abnormal behavior detection. Based on the robustness and adaptability of machine learning models, the detection and classification of intrusion data are significantly enhanced. Our presented system is validated using the CICEV2023 dataset. Simulation results demonstrate that our algorithm achieves higher efficiency and accuracy compared to existing centralized schemes.

Keywords

Intrusion Detection Systems, Abnormal Behaviour Detection, Internet of Vehicles, Distributed Learning, Machine Learning

1. Introduction

With increasing cyber threats in the era of connectivity, effective security measures within the Internet of Vehicles (IoV) are critical [1]. In an age of pervasive digital transformation, the integrity and resilience of vehicular networks are more vital than ever. The IoV integrates vehicles with EV stations, offering connectivity and convenience but also facing significant vulnerabilities to cyber attacks [2]. These vulnerabilities, if exploited, can compromise not only data but also the safety and operational efficiency of transportation systems. Traditional security techniques in the IoV have limitations, necessitating more advanced and efficient solutions [3].

Intrusion Detection Systems (IDS) play a crucial role in detecting and preventing cyber attacks in the IoV [4]. Given the rapid evolution of cyber threats, it is imperative that IDS evolve beyond traditional methods to detect both known and novel attack vectors in real time. IDS can be categorized into signature-based and abnormal behavior-based techniques, each with its strengths and weaknesses [5]. Abnormal behavior detection is essential for securing the IoV, as it can identify previously unseen attacks. Integrating machine learning techniques into abnormal behavior detection enhances the effectiveness of IDS in the IoV [6].

However, current IDS approaches face challenges such as scalability and time-consuming analysis, calling for more efficient solutions[7]. Addressing these challenges requires a paradigm shift towards distributed and intelligent security frameworks.To address these limitations, we propose a novel IDS design that leverages machine learning for efficient and accurate abnormal behavior detection. Our proposed framework is engineered to adapt dynamically to emerging threats, reduce detection latency, and enhance overall system robustness. The objectives of our approach are to overcome limitations, improve IoV security, and enhance abnormal behavior detection.

The 13th International Conference on Research in Computing at Feminine (RIF2024), 20-21 May, 2024, Constantine, Algeria
joalfarra@effat.edu.sa (J. AlFarra); leaalam@effat.edu.sa (L. Alam); naila.marir@univ-constantine2.dz (N. Marir); asarirete@effatuniversity.edu.sa (A. Sarirete)

The main contribution of this paper is the introduction of a novel IDS design that integrates distributed learning techniques and machine learning models to enhance security within the IoV. We validate our proposed system using the CICEV2023 dataset [8] and demonstrate through simulations that our algorithm achieves higher efficiency and accuracy compared to existing centralized schemes.

The paper is structured as follows: Section 2 outlines the theoretical foundations and previous works related to the proposed system. Section 3 presents the proposed system, detailing the methodology and design modules. Section 4 of the study includes a discussion on the dataset used, validation of the proposed method, and the subsequent discussion of the findings. Finally, Section 5 concludes the paper by summarizing the main findings, highlighting the contributions, and suggesting future research directions.

2. Related Works

The IoV is a rapidly evolving domain that integrates vehicles, infrastructure, and communication technologies. This integration creates a complex, interconnected environment where robust cybersecurity is paramount. Ensuring cybersecurity in the IoV is crucial, and intrusion detection plays a vital role in identifying and mitigating cyber threats. In this literature review, we systematically examine key research studies that have advanced intrusion detection techniques tailored specifically for the IoV environment.

Alladi et al. [9] proposed a deep learning-based classification framework that effectively identified misbehaving vehicles, enhancing cybersecurity in the IoV. They utilized deep learning techniques to classify and detect anomalous behaviors. In another paper, Alladi et al. [10] proposed an AI-based intrusion detection architecture for the IoV. Their approach, which deployed Deep Learning Engines (DLEs) on Multi-access Edge Computing (MEC) servers, enabled real-time identification and classification of cyber-attacks, offering robust protection against threats specific to vehicular networks.

Yang et al. conducted multiple studies focusing on intrusion detection in the IoV. In their study [11], they developed the LCCDE framework, an ensemble IDS that achieved high accuracy in detecting attacks in intra-vehicle and external networks. In another study [12], they proposed a multitiered hybrid IDS combining signature-based and anomaly-based approaches, demonstrating high accuracy and real-time efficiency. Additionally, they developed an intelligent IDS for AVs and the IoV using tree-structure machine learning models in their study [13], accurately detecting and classifying various cyber-attacks to enhance transportation system safety. Yang et al.'s research contributes to effective intrusion detection strategies in the IoV.

Laisen et al. [14] designed a data-driven IDS for the IoV using Convolutional Neural Networks (CNNs) to analyze RSU link load behaviors and achieve effective intrusion detection. Ullah et al. [15] proposed a hybrid deep learning model leveraging LSTM and GRU architectures, achieving high accuracy in detecting DDoS attacks and car hacks. Li et al. [16] addressed the challenge of updating intrusion detection models in the IoV by proposing model update schemes utilizing labeled and unlabeled data. Ahmed et al. [17] developed a deep learning-based IDS for the IoV's CAN, detecting malicious attacks and ensuring trust, privacy, and data integrity. These studies contribute to IoV security by employing deep learning techniques for intrusion detection.

The limitations of existing intrusion detection techniques and systems in the IoV, including scalability issues, big datasets, and computational requirements, can be mitigated through the use of distributed intrusion detection systems. By distributing the detection mechanisms across multiple nodes or devices within the IoV infrastructure, these systems can handle large-scale deployments, effectively process and analyze massive amounts of data, and leverage the collective processing power of multiple nodes. This approach enables improved scalability, efficient handling of big datasets, and the ability to meet the computational demands of intrusion detection in the IoV environment.

Our proposed approach addresses the research gaps in IDS for the IoV by leveraging distributed learning techniques and addressing challenges associated with centralized machine learning approaches. We introduce a novel distributed system for IDS design in the IoV, improving scalability, time-consuming

analysis, and real-time processing. By utilizing different machine learning models, we enhance the detection and classification of intrusion data, ensuring a more effective defense against potential attacks in the dynamic IoV environment. Additionally, our approach prioritizes trust and data integrity in the IoV ecosystem, contributing to a comprehensive and accurate IDS solution.

Despite the significant advancements highlighted in the literature, substantial challenges remain. Future research must focus on refining the integration of distributed architectures with advanced machine learning models to not only improve scalability and detection accuracy but also to minimize computational overhead. Moreover, real-world validations and longitudinal studies are essential to ensure that these systems can adapt to evolving threat landscapes and dynamic network conditions. Addressing these challenges will be crucial for translating theoretical advances into practical, resilient security solutions that can safeguard the IoV ecosystem in an increasingly interconnected world.

3. Proposed System

Our objective is to advance intrusion detection within the IoV by integrating theoretical foundations with practical applications. To achieve this, we propose a cluster-based distributed system, which includes master and worker nodes. In our methodology, the cluster is divided into these nodes, with the workers responsible for training the intrusion detection model using local data. Model updates, such as parameters or gradients, are subsequently transmitted to the master node to form the global model. This collaborative approach enhances security and communication system reliability, bolstering the IoV's scalability.

3.1. Architectural Framework

Figure 1 illustrates the key stages involved in developing a DL IDS for the IoV. This figure emphasizes the interconnectedness and interdependencies of the different layers, highlighting the holistic approach necessary to address the security challenges arising from increased vehicle connectivity.

1. Data Collection Layer:

The data collection layer plays a pivotal role in our approach, involving the gathering of pertinent data from various sources within the IoV environment, such as sensors, cameras, and communication networks. In line with the primary focus of our study on identifying DoS and DDoS attacks targeting EV charging stations, we utilize the CICEV2023 DDoS attack dataset [8]. This dataset provides a realistic and challenging environment for testing our proposed IDS. Once collected, the data is stored securely in a distributed storage system, such as the Hadoop Distributed File System (HDFS). This ensures that the data is accessible and protected against unauthorized access or tampering. Measures are also taken to ensure the quality and integrity of the stored data, including data validation and error detection mechanisms.

2. Data Pre-processing Layer:

The data pre-processing layer ensures that the collected data from the CICEV2023 DDoS attack dataset is clean, well-structured, and compatible for analysis within our DL IDS. The pre-processing steps include data cleaning, where irrelevant or noisy data is removed and missing values are handled appropriately. Encoding techniques are applied to categorical variables to convert them into numerical format, ensuring compatibility with machine learning algorithms. Additionally, time series management techniques are employed to organize the data into time-stamped sequences, allowing for the detection of temporal patterns associated with DoS/DDoS attacks. These pre-processing steps are essential for preparing the data for input into our DL IDS, ensuring that it can effectively detect and respond to security threats in the IoV environment. Notably, these pre-processing steps are efficiently carried out using Apache Spark, a distributed computing framework, enabling scalability and high-performance processing of large-scale datasets.

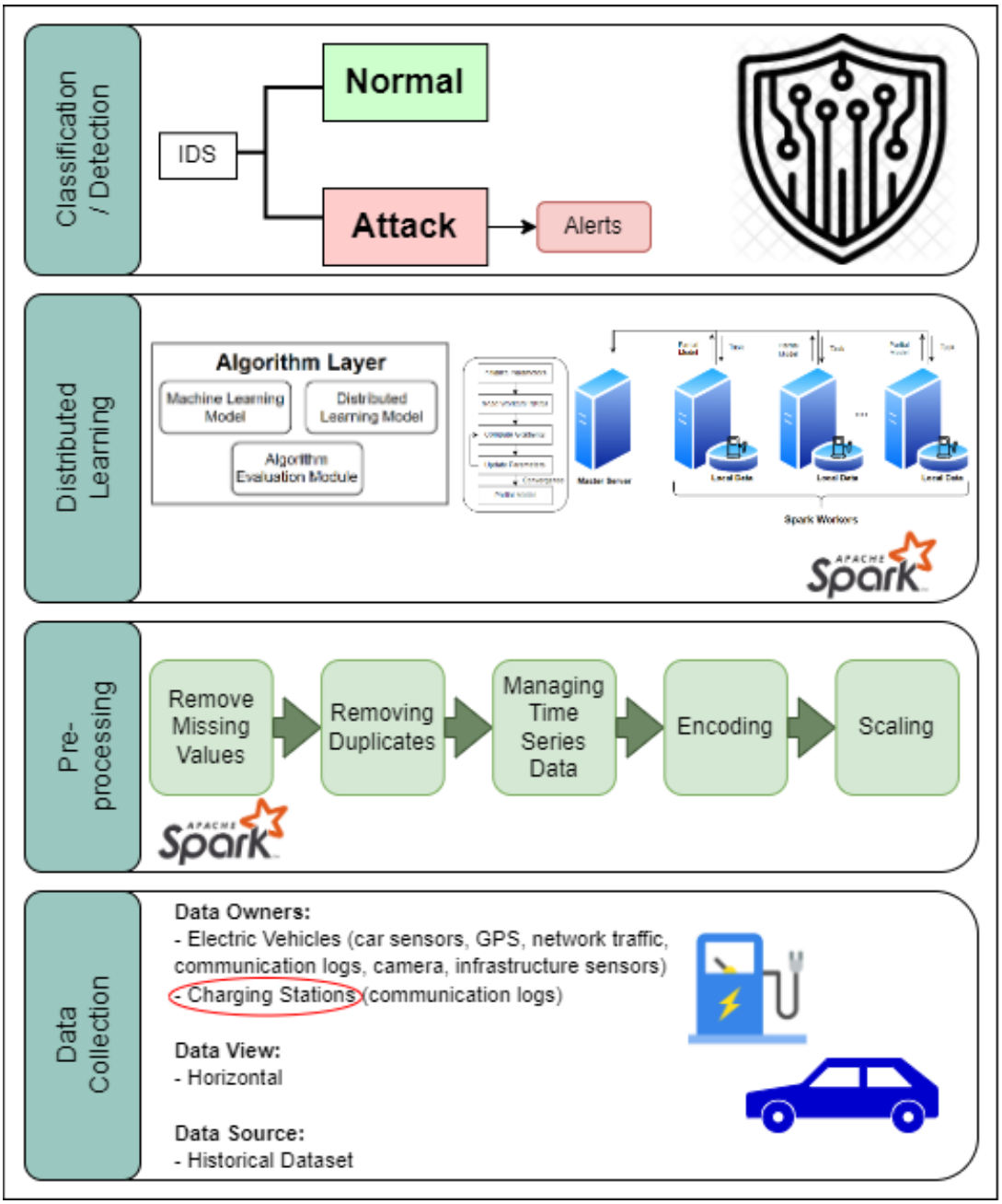

Figure 1: DL IDS Process

3. Distributed Learning Layer:

The Distributed Learning Layer in our approach involves the collaborative training of machine learning models by distributed entities within a Spark-based cluster. This process utilizes the parallel processing capabilities of Apache Spark to train the models efficiently and effectively. In a Spark-based cluster, the training data is distributed across multiple nodes, with each node being responsible for processing a subset of the data. The Spark framework manages the distribution of the data and the coordination of the training process across the cluster. During training, each node independently updates its portion of the model using local data. These updates, such as model parameters or gradients, are then aggregated at a master node to create a global model. This collaborative approach to training allows for the efficient use of computational resources and enables the training of large-scale machine learning models on big data. Furthermore, the distributed nature of the training process enhances the scalability of our approach, allowing for the training of complex models on large datasets. Additionally, by leveraging

Spark's fault-tolerant capabilities, our approach ensures that the training process is robust and reliable, even in the presence of node failures or network issues.

4. Detection Layer:
The Detection Layer in our approach is where the trained machine learning models are applied to detect DoS and DDoS attacks targeting EV charging stations within the IoV environment. This layer involves the following key components and processes:

- Model Application: The trained machine learning models, which have been developed and optimized during the training phase, are applied to the incoming data streams from EV charging stations. These models analyze the data in real-time to detect patterns and anomalies indicative of DoS/DDoS attacks.
- Real-time Monitoring: The Detection Layer continuously monitors the data streams from EV charging stations for any suspicious activity or deviations from normal behavior. This real-time monitoring allows for the immediate detection and response to potential attacks, minimizing the impact on the IoV environment.
- Alert Generation: When a potential attack is detected, the Detection Layer generates alerts or notifications to alert the system administrators. These alerts provide details about the detected attack, including the type of attack and the affected EV charging stations, enabling timely and effective response actions.

3.2. Distributed machine learning-based IDS for EV charging stations in IoV

Algorithm 1 provides an overview of the distributed training process for a machine learning-based IDS tailored for EV charging stations within the IoV. This subsection delves deeper into the workings of this algorithm, focusing on its implementation and the intricacies of training a model that can effectively detect and mitigate attacks in real-time.

4. Evaluation and Experimental Results

4.1. Dataset

In the data collection phase, we obtained a comprehensive dataset named CICEV2023, which offers critical insights into the vulnerabilities inherent in EV charging infrastructure. This dataset was meticulously curated to capture a wide range of attack scenarios, making it an invaluable resource for developing and validating our IDS. To develop an IDS, we performed extensive preprocessing to ensure the dataset's quality and compatibility with machine learning algorithms.

The dataset used for analyzing IDS performance encompasses various attack characteristics and scenarios. It includes both full-scale attacks, where multiple charging stations are simultaneously targeted, and more subtle random attacks that simulate targeted breaches. Additionally, attacks are categorized based on their statistical distribution: Gaussian attacks, which mimic normal operational patterns, and non-Gaussian attacks, which deviate significantly from typical behavior. Within this dataset, four primary attack scenarios were examined: the correct ID of the EV, wrong ID of the EV, wrong timestamp of the EV, and wrong timestamp of the charging station (CS). These scenarios were carefully selected to reflect both common vulnerabilities and critical edge cases within the IoV ecosystem.

The collected data underwent a series of rigorous preprocessing steps, which include combining the Normal and Attack data frames into a single, unified data frame, analyzing the data frame's structure and intrinsic properties, removing duplicate columns to eliminate redundancy, converting time columns to datetime objects, adjusting for timezone differences, and calculating relevant durations. This comprehensive preprocessing pipeline ensures that the dataset is robust and well-suited for

Algorithm 1 Distributed Training Process for EV Charging Station IDS

Initialization:
Initialize a global model W at the master node (server).
Training Process:
for $t = 1$ **to** T **do**
  Client Selection:
  Randomly select a subset of K charging stations from the network to participate in the training process.
  Client Model Update:
  for each selected charging station k **do**
    Download the current global model W from the master node.
    Perform local model training on the charging station's local dataset D_k to obtain a new local model W_k'.
  end for
  Model Aggregation:
  Aggregate the local models W_k' from all selected charging stations:
  $W_t = \frac{1}{K} \sum_{k=1}^{K} W_k'$
  Model Update:
  Upload the aggregated model W_t to the master node.
  The master node updates the global model:
  $W = W + \eta \cdot (W_t - W)$, where η is the learning rate.
end for
Intrusion Detection and Notification:
for each new data sample x received from a charging station **do**
  Use the global model W to predict whether the sample is normal or an attack.
  if an attack is detected **then**
    Notify nearby vehicles in the network to avoid using the compromised charging station.
  end if
end for
Termination:
Repeat the training process for a predefined number of rounds T or until convergence criteria are met.

advanced machine learning analyses, ultimately enhancing the reliability of our intrusion detection system.

4.2. Validation of the Proposed Method

To validate the effectiveness of the proposed system, a series of experiments and evaluations were conducted. The dataset was divided into training and testing sets to ensure representative and unbiased training of the IDS. The training set was used to train the 20 machine learning models, while the testing set was employed to evaluate the performance of the IDS.

Various performance metrics were employed to assess the effectiveness of the proposed system. These metrics included accuracy, precision, recall, and F1-score, as shown in Figure 2. Accuracy measures the overall correctness of the IDS in classifying attacks and normal behavior. Precision examines the proportion of correctly classified attacks out of all detected attacks, while recall calculates the proportion of correctly classified attacks out of all actual attacks. The F1-score combines precision and recall to provide an overall measure of the IDS's performance.

The bar chart in Figure 3 provides a visual representation of the accuracy levels achieved by several models in intrusion detection. It reveals variations in performance among different models, with some, such as LGBM, XGB, DecisionTree, RandomForest, and ExtraTrees, demonstrating high accuracy.

Model	Accuracy	Precision	Recall	F1-Score
LGBMClassifier	99.53502343	0.994955126	0.990621245	0.992770009
RidgeClassifierCV	80.2006488	0.874320592	0.51158293	0.467700006
XGBClassifier	99.85582122	0.998874192	0.996661159	0.997762981
QuadraticDiscriminantAnalysis	79.06403941	0.634724627	0.555944324	0.558349947
CalibratedClassifierCV	80.41211102	0.882971147	0.516804552	0.478081725
BernoulliNB	79.7512916	0.398756458	0.5	0.443675764
BaggingClassifier	99.84741079	0.998933245	0.996342803	0.997631593
LogisticRegression	80.36405142	0.881439823	0.51561782	0.475739636
NearestCentroid	59.77411991	0.529318839	0.542873023	0.513933628
SVC	94.63174336	0.955500211	0.876517586	0.909298965
LinearSVC	80.19824582	0.874184398	0.511523593	0.467580885
KNeighborsClassifier	99.56025472	0.996938056	0.989451297	0.993140715
GaussianNB	78.08602667	0.587118706	0.531418016	0.522735196
Perceptron	80.06127598	0.859502413	0.508207813	0.460914487
SGDClassifier	79.7512916	0.398756458	0.5	0.443675764
DecisionTreeClassifier	99.79814971	0.998625335	0.995126402	0.996864128
RandomForestClassifier	99.90388081	0.999264417	0.997759349	0.998509714
MLPClassifier	98.45848853	0.982855664	0.969107488	0.975793098
ExtraTreesClassifier	99.90868677	0.999316783	0.997855886	0.998584291
AdaBoostClassifier	86.04950138	0.901896224	0.661962726	0.703614892

Figure 2: Results of the ML Models

Table 1
Accuarcy Comparison

Paper	Machine Learning Algorithm	Accuracy
[11]	LCCDE	99.81%
[13]	Stacking	99.86%
[17]	Deep Learning	96%
Proposed Solution	Distributed Learning	99.91%

However, other models, like Ridge, QDA, BernoulliNB, and GaussianNB, exhibit relatively lower accuracy. This chart provides valuable insights into the relative performance of different models, assisting in the identification of suitable options for effective intrusion detection tasks.

The results obtained from the validation process serve as a strong validation of the proposed system's capabilities and contribute to its credibility and reliability in real-world intrusion detection scenarios.

4.3. Discussion

Our proposed system takes a novel approach by leveraging distributed learning techniques to design an efficient Intrusion Detection System (IDS) for EV charging stations within the IoV.

In comparing our results to existing solutions, we have achieved notably higher levels of efficiency and accuracy in detecting and classifying intrusion data. Our system leveraged a diverse set of machine learning models, which contributed to the enhanced performance of our intrusion detection system (IDS). The accuracy of our IDS ranged from 59.77% to 99.91%, with an average accuracy of 89.96%. To provide a comprehensive overview and facilitate a more intuitive understanding of the advancements made, we have included Table 1 below. This table showcases the accuracies reported for the existing approaches, underscoring the significant progress achieved in the field and the effectiveness of our proposed methodology.

While the specific metrics for each existing solution were not provided in the literature review,

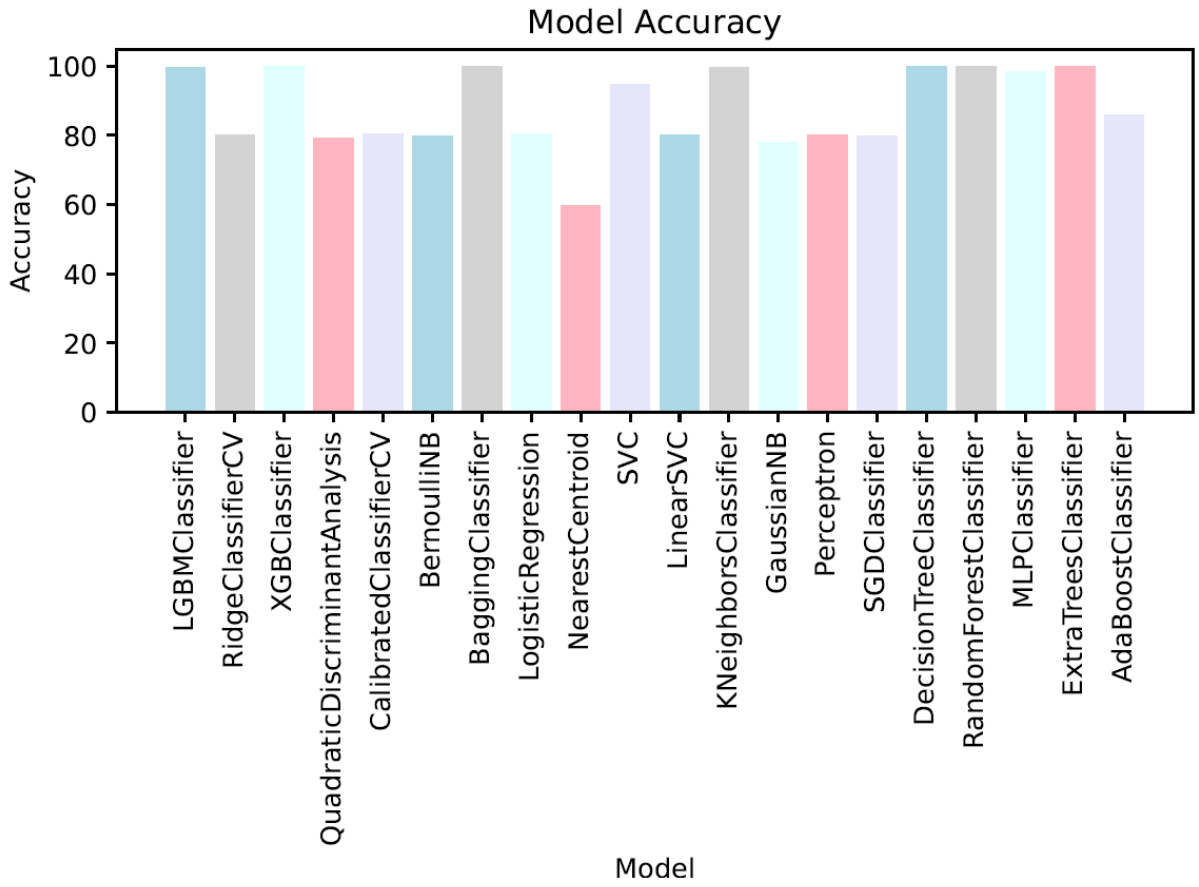

Figure 3: Comparing Model Accuracies for Intrusion Detection

our system achieved high accuracy rates while maintaining a balance between precision and recall. This indicates that our IDS has a low false positive rate (precision) while effectively detecting a high proportion of actual attacks (recall). The F1-score, which combines precision and recall into a single measure, further demonstrates the overall performance of our system.

By leveraging distributed learning, our system overcomes the limitations of centralized machine learning approaches used in some existing solutions. This allows for efficient and real-time processing of data, enhancing the effectiveness of abnormal behavior detection within EV charging stations. Our results demonstrate the effectiveness of our system in enhancing the security of EV charging stations within the interconnected network of the IoV.

5. Conclusion

In conclusion, this study presents a robust distributed learning-based IDS framework that significantly enhances the security of EV stations within the IoV. By addressing the limitations of centralized machine learning approaches, the proposed framework improves the effectiveness of abnormal behavior detection and intrusion data classification. The presented system, validated using the CICEV2023 dataset, achieves higher efficiency and accuracy compared to existing centralized schemes. Additionally, the framework demonstrates its efficacy in detecting and mitigating DoS and DDoS attacks through the utilization of the CICEV2023 DDoS Attack Dataset and Apache Spark for data preprocessing and distributed training.

This research contributes substantially to the field of IoV security by offering a comprehensive solution that not only fortifies EV charging stations but also reinforces the overall integrity of vehicular networks. By leveraging distributed learning, our approach enhances scalability and ensures real-time responsiveness, which are critical for defending against rapidly evolving cyber threats in a highly

interconnected environment.

Looking ahead, future research should focus on further optimizing the proposed framework to address emerging challenges. Potential directions include integrating advanced encryption techniques for secure data transmission, as well as implementing privacy-preserving strategies such as federated learning and differential privacy to safeguard sensitive information against intrusion attacks. Longitudinal studies and real-world deployments will be essential to validate the robustness and adaptability of the framework under diverse operational conditions. These efforts will be pivotal in translating theoretical advancements into practical, industry-ready solutions that can sustain the dynamic and complex nature of the IoV ecosystem.

Overall, our work paves the way for the development of next-generation IDS frameworks that are not only effective in thwarting cyber threats but also scalable, resilient, and adaptable to the evolving demands of modern vehicular networks.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT for rephrasing and improving clarity of certain paragraphs. All content generated or suggested by these tools was critically reviewed and edited by the authors. The authors affirm full responsibility for the accuracy, originality, and integrity of the paper.

References

[1] H. Taslimasa, S. Dadkhah, E. C. P. Neto, P. Xiong, S. Ray, A. A. Ghorbani, Security issues in internet of vehicles (iov): A comprehensive survey, Internet of Things (2023) 100809.

[2] S. M. Karim, A. Habbal, S. A. Chaudhry, A. Irshad, et al., Architecture, protocols, and security in iov: Taxonomy, analysis, challenges, and solutions, Security and Communication Networks 2022 (2022).

[3] P. K. Sadhu, V. P. Yanambaka, A. Abdelgawad, Internet of things: Security and solutions survey, Sensors 22 (2022) 7433.

[4] D. Man, F. Zeng, J. Lv, S. Xuan, W. Yang, M. Guizani, Ai-based intrusion detection for intelligence internet of vehicles, IEEE Consumer Electronics Magazine 12 (2021) 109–116.

[5] M. Ozkan-Okay, R. Samet, Ö. Aslan, D. Gupta, A comprehensive systematic literature review on intrusion detection systems, IEEE Access 9 (2021) 157727–157760.

[6] M. Houmer, M. Ouaissa, M. Ouaissa, S. Eddamiri, Applying machine learning algorithms to improve intrusion detection system in iov, Artificial Intelligence of Things in Smart Environments: Applications in Transportation and Logistics (2022) 35.

[7] A. Khraisat, A. Alazab, A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges, Cybersecurity 4 (2021) 1–27.

[8] University of new brunswick est.1785, ???? URL: https://www.unb.ca/cic/datasets/cicev2023.html.

[9] T. Alladi, V. Kohli, V. Chamola, F. R. Yu, Securing the internet of vehicles: A deep learning-based classification framework, IEEE networking letters 3 (2021) 94–97.

[10] T. Alladi, V. Kohli, V. Chamola, F. R. Yu, M. Guizani, Artificial intelligence (ai)-empowered intrusion detection architecture for the internet of vehicles, IEEE Wireless Communications 28 (2021) 144–149.

[11] L. Yang, A. Shami, G. Stevens, S. De Rusett, Lccde: A decision-based ensemble framework for intrusion detection in the internet of vehicles, in: GLOBECOM 2022-2022 IEEE Global Communications Conference, IEEE, 2022, pp. 3545–3550.

[12] L. Yang, A. Moubayed, A. Shami, Mth-ids: A multitiered hybrid intrusion detection system for internet of vehicles, IEEE Internet of Things Journal 9 (2021) 616–632.

[13] L. Yang, A. Moubayed, I. Hamieh, A. Shami, Tree-based intelligent intrusion detection system in internet of vehicles, in: 2019 IEEE global communications conference (GLOBECOM), IEEE, 2019, pp. 1–6.
[14] L. Nie, Z. Ning, X. Wang, X. Hu, J. Cheng, Y. Li, Data-driven intrusion detection for intelligent internet of vehicles: A deep convolutional neural network-based method, IEEE Transactions on Network Science and Engineering 7 (2020) 2219–2230.
[15] S. Ullah, M. A. Khan, J. Ahmad, S. S. Jamal, Z. e Huma, M. T. Hassan, N. Pitropakis, Arshad, W. J. Buchanan, Hdl-ids: a hybrid deep learning architecture for intrusion detection in the internet of vehicles, Sensors 22 (2022) 1340.
[16] X. Li, Z. Hu, M. Xu, Y. Wang, J. Ma, Transfer learning based intrusion detection scheme for internet of vehicles, Information Sciences 547 (2021) 119–135.
[17] I. Ahmed, G. Jeon, A. Ahmad, Deep learning-based intrusion detection system for internet of vehicles, IEEE Consumer Electronics Magazine 12 (2021) 117–123.