International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Hybrid Machine Learning Framework for Cyber Attack Detection and Attribution in IoT-Enabled Cyber-Physical Systems

G. Richitha, R. Vamsi, CH. Lakshmi Prasanna, K.N. Abdul Shukoor

B-Tech in Electronics and Communication Engineering, Department of Electronics and Communication Engineering, Sanketika Vidya Parishad Engineering College

Ms. R. Geethika

Assistant Professor, Department of Electronics and Communication Engineering, Sanketika Vidya Parishad Engineering College

---

Abstract -
Securing Internet-of-Things (IoT)-enabled Cyber-Physical Systems (CPS) is a challenging task due to increased connectivity and system complexity. Traditional security solutions designed for IT/OT environments are insufficient for CPS. This paper proposes a **two-stage ensemble deep learning framework** for cyber-attack detection and attribution in Industrial Control Systems (ICS).

In the first stage, an ensemble representation learning model combined with a Decision Tree classifier detects cyber-attacks in imbalanced datasets. In the second stage, a Deep Neural Network (DNN) performs attack attribution using one-vs-all classifiers. The model is evaluated using the SWAT dataset and demonstrates improved accuracy, recall, and F-measure compared to existing techniques.

Key Words : IoT, Cyber Physical Systems, Cyber Security, Deep Learning, Autoencoder, Decision Tree, DNN, Attack Detection, Attack Attribution.

1. INTRODUCTION

The integration of IoT devices into Cyber-Physical Systems (CPS) has significantly improved automation in industries such as smart grids, healthcare, and manufacturing. However, this integration also increases vulnerability to cyber-attacks. Industrial Control Systems (ICS), including SCADA and PLC systems, are critical infrastructures that require high security. Unlike IT systems, ICS prioritizes **availability over confidentiality**, making cyber-attacks more dangerous. Several real-world attacks such as **Stuxnet**, **BlackEnergy**, and gas pipeline attacks highlight the need for robust detection systems. Traditional methods like signature-based detection are ineffective against unknown attacks.To address these challenges, this work proposes a **hybrid deep learning-based attack detection and attribution system** capable of detecting both known and unknown attacks.

2. LITERATURE SURVEY

The rapid growth of Internet of Things (IoT) technologies and their integration into Cyber-Physical Systems (CPS) have introduced significant security challenges. Industrial Control Systems (ICS), which form a critical part of CPS, are highly vulnerable to cyber-attacks due to their dependence on networked communication and real-time operations. Traditional intrusion detection systems (IDS), designed for conventional IT environments, are not suitable for CPS due to differences in system behaviour and data characteristics. Early approaches to intrusion detection utilized Machine Learning (ML) algorithms such as Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Decision Trees (DT), and Random Forest (RF). These methods provided acceptable performance in detecting known attacks; however, they suffer from high false positive rates and poor performance on imbalanced datasets. In ICS environments, where attack instances are rare compared to normal data, these models tend to misclassify critical attack samples. To overcome these limitations, Deep Learning (DL) techniques have been widely adopted. Models such as Deep Neural Networks (DNN) have shown improved accuracy due to their

ability to learn complex patterns from large datasets. Additionally, autoencoder-based models have been used for anomaly detection by learning normal system behavior and identifying deviations as potential attacks. These models are effective in detecting unknown attacks but have limitations in classifying different attack types. Recent research has focused on hybrid and ensemble approaches that combine multiple models to improve detection performance.For example,integrating autoencoders with classification algorithms enhances both feature extraction and classification accuracy. Despite these advancements, challenges such as detecting zero-day attacks, handling imbalanced datasets, and performing attack attribution remain unresolved.

3. METHODOLOGY

This paper proposes a **two-stage ensemble deep learning framework** for detecting and attributing cyber-attacks in IoT-enabled Cyber-Physical Systems (CPS). The methodology is designed to handle imbalanced datasets, improve detection accuracy, and classify different types of attacks efficiently. The overall process consists of data preprocessing, feature extraction, dimensionality reduction, attack detection, and attack attribution.

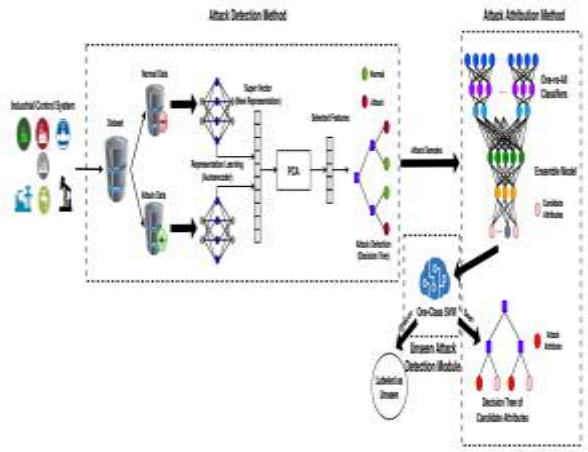

Fig1: Attack Detection & Attribution Framework

Dataset Preparation: The input dataset is divided into normal and attack data collected from Industrial Control Systems.

Representation Learning: Autoencoders are used to extract deep features from both normal and attack data, forming a super-vector representation.

Dimensionality Reduction: Principal Component Analysis (PCA) is applied to select important features and reduce complexity.

Attack Detection: A Decision Tree classifier identifies normal and attack data, while a One-Class SVM detects unseen attacks.

Attack Attribution: Detected attack samples are classified using a Deep Neural Network with a one-vs-all approach.

Ensemble Learning: Multiple classifiers are combined to improve accuracy and provide reliable candidate attributes.

Final Output: The system labels attacks and provides accurate detection and attribution results for monitoring and decision-making.

3.1 Data Preprocessing

The dataset used in this work is the SWAT (Secure Water Treatment) dataset, which includes both normal and attack scenarios. Initially, preprocessing is performed to enhance data quality. Missing values are handled, and irrelevant features are removed. The dataset is then normalized using min-max scaling to bring all features into a uniform range. This step improves model performance and ensures faster convergence during training.

3.2 Feature Extraction using Autoencoder

In the first stage, feature extraction is carried out using stacked autoencoders. Two separate autoencoder models are trained:

- One using normal data
- One using attack data

These models learn compressed representations of input data by minimizing reconstruction error. The encoded features from both models are combined to form a unified feature vector, which enhances the distinction between normal and malicious data.

Fig2: Training of Autoencoder Algorithmn

3.3 Classification using Decision Tree:

Decision Tree is a supervised learning algorithm widely used for classification tasks. It works by splitting the data into subsets based on feature values, forming a tree-like structure of decision rules. Each internal node represents a feature condition, and each leaf node represents a class label. In the proposed system, the Decision Tree classifier is used to differentiate between normal and attack data due to its simplicity, interpretability, and ability to handle non-linear relationships efficiently.

Fig3: Training of Decision Tree Algorithmn

3.4 Attack Attribution using Deep Neural Network (DNN):

Deep Neural Networks are advanced learning models capable of capturing complex and non-linear relationships in large datasets. In this framework, the DNN is employed for attack attribution, where it classifies detected attacks into specific categories. The model uses a one-vs-all classification approach and a Softmax function to determine the final class label. This enables accurate identification of different attack types, thereby improving the reliability and effectiveness of the system.

Fig4: Training of DNN Algorithmn

4. RESULTS

4.1 Attack Detection & Attack Attribution:

The proposed system performs cyber-attack detection and attribution using a two-stage intelligent framework designed for IoT-enabled Cyber Physical Systems. In the first stage, a Decision Tree model is employed to analyze the preprocessed and feature-engineered data to effectively distinguish between normal and malicious activities. This enables fast and reliable detection of potential cyber-attacks in real-time environments. In the second stage, the detected attack samples are forwarded to a Deep Neural Network (DNN), which performs detailed analysis to accurately classify the type of attack, such as Denial of Service (DoS), NMRI, CMRI, MSCI, MPCI, and MFCI. Additionally, mechanisms such as One-Class SVM can be incorporated to identify previously unseen or unknown attack patterns, enhancing the robustness of the system. This integrated approach ensures both efficient detection and precise attribution, thereby significantly improving the security and resilience of IoT-based industrial systems.

Fig5: Detection & Attribution of Cyber Attacks

4.2 Comparision Graph:

In the below graph x-axis represents algorithms names and y-axis represents different metric values such as precision, recall, accuracy and FSCORE with different colour bars and in all algorithms DNN got high.

Fig6: Comparision Graph of Different Algorithmns

4.3 Comparision Table:

In the below table we can see algorithm names and its metrics values such as accuracy and precision and other.

Fig7: Comparision Table

5.CONCLUSION AND FUTURE SCOPE

This work presents an effective two-stage framework for the detection and attribution of cyber-attacks in IoT-enabled Cyber Physical Systems. The proposed approach integrates a Decision Tree model for rapid and accurate attack detection with a Deep Neural Network for precise classification of attack types. The use of advanced techniques such as autoencoder-based feature extraction and Principal Component Analysis enhances the quality of input data and improves overall system performance. Furthermore, the incorporation of mechanisms to identify unseen attacks increases the robustness of the framework against evolving cyber threats. Experimental evaluation on the SWaT dataset demonstrates that the proposed model achieves superior performance in terms of accuracy, precision, recall, and F-score compared to existing methods. Overall, the system provides a reliable and efficient solution for enhancing the security and resilience of industrial IoT environments.

Future work can focus on extending the proposed framework for real-time deployment in large-scale industrial environments to enable continuous monitoring and faster response to cyber threats. The integration of advanced deep learning models such as LSTM and CNN can further improve the system's ability to capture temporal patterns and enhance detection accuracy. Additionally, incorporating predictive analytics can enable early identification of potential attacks before they fully manifest. The development of automated response mechanisms can help in immediate mitigation of detected attacks without human intervention. Furthermore, deploying the system on cloud platforms can improve scalability and accessibility, while the use of more diverse and large-scale datasets can enhance the model's generalization capability. These improvements will contribute to building more intelligent, adaptive, and secure IoT-enabled cyber-physical systems.

REFERENCES

[1] Y. Zhang, L. Wang, W. Sun, R. Green II, and M. Alam, "Distributed intrusion detection system in a multi-layer network architecture of smart grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 796–808, 2011.

[2] R. Mitchell and I. R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Computing Surveys*, vol. 46, no. 4, pp. 1–29, 2014.

[3] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *Proceedings of the 3rd Conference on Hot Topics in Security (HotSec)*, 2008, pp. 1–6.

[4] S. Adepu and A. Mathur, "Distributed attack detection in a water treatment plant: Method and case study," *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 1, pp. 1–15, 2020.

[5] J. Goh, S. Adepu, M. Tan, and Z. S. Lee, "Anomaly detection in cyber physical systems using recurrent neural networks," in *IEEE International Symposium on High Assurance Systems Engineering*, 2017, pp. 140–145.