Project Proposal

On

Hybrid ML-based DDoS Attack Detection in Industry 4.0 Cyber-Physical Production Systems (CPPSs)

Submitted By:
ZM Abul Kasem Noyon
20CSE015
Department of Computer Science And Engineering
University of Barisal

Submitted To:
Md. Rashid Al Asif
Assistant Professor
Department of Computer Science And Engineering
University of Barisal

Submission Date: 10 September 2025

Abstract

With the advent of Industry 4.0, Cyber-Physical Production Systems (CPPSs) have revolutionized industrial operations, offering immense benefits in terms of efficiency, automation, and real-time data analysis. However, these advancements also introduce new cybersecurity threats, particularly Distributed Denial-of-Service (DDoS) attacks, which can severely impact the availability and performance of critical systems. Traditional intrusion detection systems (IDS) have limitations when it comes to detecting sophisticated DDoS attacks in real-time. This research proposes a Hybrid Machine Learning (ML)-based approach for detecting DDoS attacks in Industry 4.0 CPPSs. By combining supervised, unsupervised, and semi-supervised learning models, the study aims to develop a more accurate and robust detection system that minimizes false positives while maintaining high detection accuracy. The proposed model will be evaluated using CICIDS 2019 and other public datasets, and compared with traditional single-model IDS approaches. The results are expected to enhance the security of Industry 4.0 systems by providing real-time, scalable, and reliable DDoS detection.

Keywords: : Industry 4.0, Cyber-Physical Production Systems (CPPSs), DDoS Detection, Hybrid Machine Learning, Anomaly Detection, Supervised Learning, Unsupervised Learning, Semi-supervised Learning, Ensemble Learning, Real-time Detection, Cybersecurity, Network Traffic Analysis, Data Science, IoT Security, Public Datasets

1. Introduction

The Fourth Industrial Revolution (Industry 4.0) marks a revolutionary shift in the way industries operate, integrating Cyber-Physical Production Systems (CPPSs) with cutting-edge technologies such as the Internet of Things (IoT), cloud computing, big data analytics, and artificial intelligence (AI). This convergence of physical and digital systems has enabled industries to unlock new levels of automation, real-time decision-making, and operational efficiency. Furthermore, it has opened doors to customized services, allowing companies to tailor production processes and services to meet specific consumer demands in real-time.

As organizations embrace Industry 4.0, they increasingly rely on interconnected systems and devices that communicate and exchange data seamlessly across various platforms. While these advancements offer significant advantages, they also introduce new vulnerabilities. The growing interconnectivity between machines, devices, and networks creates multiple points of entry for malicious actors to exploit. As a result, the cybersecurity of CPPSs has become a critical concern, as these systems are vulnerable to a range of cyber threats that can compromise the integrity, availability, and confidentiality of industrial operations.

One of the most severe and prevalent cyber threats facing CPPSs today is Distributed Denial of Service (DDoS) attacks. These attacks involve overwhelming the network infrastructure with massive amounts of traffic, ultimately flooding system resources and rendering them unavailable. As a result, production lines can come to a halt, disrupting operations and potentially causing significant financial losses. The challenge of detecting these attacks lies in their distributed nature, where attackers use multiple compromised machines to launch coordinated assaults. This large-scale attack mechanism makes it difficult for traditional intrusion detection systems (IDS) to quickly and accurately identify and mitigate the threat.

Traditional IDS solutions, particularly those that rely on signature-based detection, are often inadequate in identifying DDoS attacks, especially zero-day attacks or new attack variants. Signature-based systems are designed to recognize patterns from known attacks, but they fail when dealing with novel or evolving attack strategies. In addition, these systems often struggle with handling the scale and complexity of real-time attack detection in modern industrial systems. This gap highlights the urgent need for more advanced and dynamic anomaly detection systems, powered by Machine Learning (ML) algorithms, which can detect unknown attacks and adapt to new attack patterns.

In this research, we propose a Hybrid ML-based approach for detecting DDoS attacks in Industry 4.0 CPPSs. The hybrid model will combine the strengths of supervised, unsupervised, and semi-supervised learning techniques to improve the overall detection accuracy while minimizing false positives. Supervised learning models will help classify known attack patterns, while unsupervised learning models will detect anomalies without prior labeling, and semi-supervised models will leverage both benign and a limited amount of malicious data for model refinement. By combining these techniques, the approach aims to deliver a comprehensive and real-time DDoS attack detection solution that can keep pace with the evolving landscape of cyber threats in the Industry 4.0 environment.

2. Objectives

1. To evaluate the performance of supervised, unsupervised, and semi-supervised learning models for DDoS attack detection in Industry 4.0 CPPSs.

2. To develop a Hybrid ML-based model combining multiple learning techniques (supervised, unsupervised, and semi-supervised) for enhanced DDoS attack detection.

3. To perform feature extraction from CICIDS 2019 and other public datasets, identifying key network traffic features relevant for attack detection.

4. To compare the performance of the Hybrid ML model with traditional single-model IDS approaches using accuracy, precision, recall, F1 score, and ROC-AUC.

5. To implement the Hybrid ML-based system and test its real-time performance on simulated Industry 4.0 CPPSs environments.

3. Problem Statement

Industry 4.0 has introduced significant transformations in industrial environments, where Cyber-Physical Production Systems (CPPSs) are integrated with technologies like IoT, cloud computing, and AI. These advancements have enhanced automation, efficiency, and real-time data processing, making industries more responsive and adaptable. However, with this increased connectivity, new vulnerabilities have emerged, making cybersecurity a critical concern.

Among the most severe threats to CPPSs are Distributed Denial of Service (DDoS) attacks, which flood the system with excessive traffic, rendering it unavailable and disrupting normal operations. These attacks can cause financial losses, production downtime, and significant damage to a company's reputation, especially in industries that rely on continuous, real-time processes.

Traditional intrusion detection systems (IDS), which typically use signature-based detection, struggle to keep up with new and evolving attack techniques. These systems rely on known attack patterns, making them ineffective against zero-day or previously unseen threats.

As a result, there is a growing need for more adaptive and accurate anomaly detection systems, especially those powered by Machine Learning (ML) algorithms.

While ML methods have shown promise in detecting network anomalies and intrusions, most current approaches rely on either supervised or unsupervised learning alone, each with its limitations. Supervised learning requires labeled datasets, which may not cover new attack variants, while unsupervised learning can produce high false positives. The goal of this research is to bridge this gap by proposing a Hybrid ML-based model that combines the strengths of both supervised and unsupervised learning techniques, offering better detection accuracy and minimizing false positives. This approach is expected to provide a more reliable solution for real-time DDoS attack detection in Industry 4.0 CPPSs.

4. Research Question

1. How can Hybrid Machine Learning models improve the accuracy and efficiency of DDoS attack detection in Industry 4.0 CPPSs?

2. What combination of supervised, unsupervised, and semi-supervised learning provides the best results for detecting different types of DDoS attacks?

3. How do hybrid models compare to traditional single-model approaches (e.g., Decision Trees, SVM) in terms of false positive rates, detection accuracy, and real-time performance?

4. How can feature selection and dimensionality reduction (e.g., PCA) contribute to enhancing the performance of the hybrid model?

5. How can real-time DDoS attack detection be implemented and optimized using a Hybrid ML approach in an industrial context?

5. Literature Review

The emergence of Industry 4.0 has introduced significant changes to industrial environments, primarily through the integration of Cyber-Physical Production Systems (CPPSs) with advanced technologies such as IoT, cloud computing, artificial intelligence (AI), and big data analytics. While these technologies have enabled industries to achieve higher automation, real-time decision-making, and operational efficiency, they have also created new vulnerabilities. These vulnerabilities have led to the increased risk of cyberattacks, which can severely impact the availability, integrity, and functionality of industrial systems.

Among the various cyber threats,DDoS attacks are particularly devastating. These attacks involve overwhelming the targeted system or network with excessive traffic, making it unavailable to legitimate users and disrupting normal operations. In the context of CPPSs, such attacks can lead to significant consequences, including financial losses, production downtime, and reputational damage. The impact of DDoS attacks can be especially severe in industries that rely on continuous, real-time processes, where even a short period of downtime can disrupt critical production cycles and cause substantial economic losses.

Traditional DDoS Detection Methods

Traditional methods for detecting DDoS attacks generally include signature-based detection and anomaly-based detection. Signature-based methods rely on known attack patterns to detect intrusions, but they are ineffective against zero-day attacks or new attack variants that do not match any existing signature. On the other hand, anomaly-based methods attempt to identify deviations from normal traffic patterns, but they often suffer from high false positive rates, particularly in dynamic environments where normal behavior can vary widely. Although traditional Intrusion Detection Systems (IDS) have been widely used in the past, they are increasingly being replaced or supplemented by Machine Learning (ML)-based methods due to their ability to detectunknown threats and adapt to new attack patterns.

Machine Learning for DDoS Detection

Machine Learning (ML) has gained significant traction in recent years for the detection of DDoS attacks. ML techniques can classify network traffic as either normal or malicious, and they are particularly useful for detecting unknown threats. The ability of supervised and unsupervised learning models to identify anomalies and patterns in large datasets has made them suitable for DDoS attack detection in modern industrial systems.

Supervised Learning Models

Supervised learning algorithms such as Random Forest (RF), Support Vector Machines (SVM), and K-Nearest Neighbors (K-NN) are often employed for DDoS detection. These models require labeled training data, which includes both benign and malicious traffic. The major advantage of supervised learning is its high accuracy when trained on high-quality, well-labeled data. However, obtaining large datasets with accurate labels is often challenging, particularly in real-world industrial systems, where DDoS attack data may be scarce.

Unsupervised Learning Models

Unsupervised learning algorithms, such as K-Means Clustering and Expectation-Maximization (EM), are also widely used for detecting anomalies in network traffic. These methods do not require labeled data and can detect previously unseen types of attacks. However, they often suffer from high false positive rates because they classify normal traffic as suspicious if it deviates from established patterns, which may not always be an attack. Despite this, unsupervised models are useful in environments where labeled data is scarce or unavailable.

Semi-supervised Learning Models

To bridge the gap between supervised and unsupervised models, semi-supervised learning techniques have been proposed. These methods leverage a small amount of labeled data along with a larger amount of unlabeled data to detect anomalies. For instance, Univariate Gaussian models have been used in semi-supervised settings for anomaly detection in DDoS attacks, where only a small amount of malicious data is required for model training. Semi-supervised learning strikes a balance between the need for labeled data and the flexibility to detect new attack patterns.

Hybrid Machine Learning Models

Recent research has increasingly focused on combining multiple ML techniques to overcome the limitations of individual models. Hybrid Machine Learning models, which combine supervised, unsupervised, and semi-supervised learning approaches, offer a promising solution to the challenges of DDoS detection in Industry 4.0 systems. The integration of multiple models allows the hybrid system to leverage the strengths of each individual method, improving detection accuracy while minimizing false positives.

For example, ensemble methods such as stacking, boosting, and bagging have been used to combine the outputs of multiple models, resulting in enhanced performance. Additionally, hybrid models allow for real-time detection of DDoS attacks, which is essential in critical industrial environments. These models are better equipped to handle the complexity and volume of traffic seen in Industry 4.0 CPPSs.

Limitations and Challenges

While hybrid ML-based DDoS detection models have shown promise, several challenges remain:

- **Data Imbalance**: Datasets used for training these models often contain an imbalance between normal and malicious traffic, which can lead to biased models with poor performance.

- **Feature Selection**: Identifying the right features for training models is crucial, as irrelevant or redundant features can degrade model performance.

- **Real-time Performance**: Ensuring that hybrid models can detect DDoS attacks in real-time, without introducing significant delays, is a major challenge.

- **Scalability**: Hybrid models need to handle the large volume and high velocity of data in Industry 4.0 environments while maintaining high accuracy and low latency.

Future Directions

The integration of federated learning, edge computing, and multi-model systems could further enhance the scalability and performance of DDoS detection in Industry 4.0 CPPSs. Additionally, exploring the combination of Deep Learning models with traditional machine learning techniques may improve the ability of these systems to detect even more sophisticated attack patterns. Research into explainable AI (XAI) will also be important for making these models more transparent and understandable to human operators, helping them make informed decisions in response to potential threats.

6. Methodology

Data Collection:

- The CICIDS 2019 dataset, along with other publicly available datasets such as KDD Cup 99, will be used for model training and testing. These datasets contain network traffic data, including DDoS attacks and normal traffic.

Feature Extraction:

- Network traffic features such as packet length, flow duration, inter-arrival time, and protocol type will be extracted using tools such as NetMate or Wireshark.

Model Development:

- **Supervised Learning Models:** Models such as Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Naïve Bayes will be trained using labeled data.
- **Unsupervised Learning Models:** K-Means and Expectation-Maximization (EM) will be used for anomaly detection based on unlabeled data.
- **Semi-supervised Learning Models:** The Univariate Gaussian algorithm will be implemented, leveraging mostly benign data with a few malicious instances for model training.

Hybrid Model:

- The Hybrid ML model will integrate the outputs of the supervised, unsupervised, and semi-supervised models to improve detection accuracy and minimize false positives. Techniques such as stacking and boosting will be employed to combine the models.

Model Evaluation:

- The models will be evaluated using cross-validation, and performance metrics such as accuracy, precision, recall,F1 score, and ROC-AUC will be used for assessment.
- Real-time detection will be tested using streaming data to simulate real-world conditions.

Sequence Diagram:

Below is a simple sequence diagram outlining the methodology flow for DDoS attack detection:

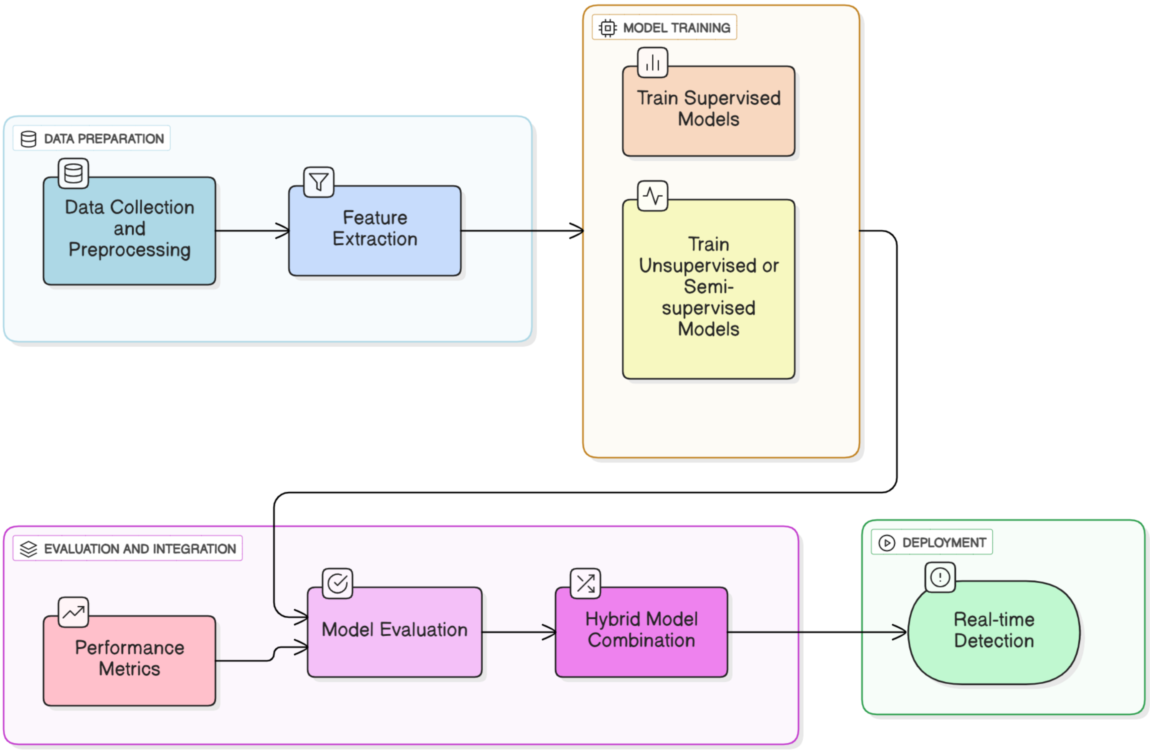

Figure 1: Sequence Diagram for DDoS Attack Detection

7. Expected Outcomes

The expected outcomes of this project are both technical and practical, aiming to strengthen the cybersecurity posture of Industry 4.0 CPPSs:

1. Improved Detection Accuracy: The hybrid model is expected to outperform single-model approaches in terms of detection accuracy and minimizing false positives.

2. Real-time Detection: The proposed model will demonstrate its ability to detect DDoS attacks in real-time in simulated Industry 4.0 environments.

3. Enhanced Performance: By combining supervised, unsupervised, and semi-supervised models, the hybrid approach will provide a more robust detection system for complex attack patterns.

4. Scalability: The proposed system will be scalable to handle large volumes of traffic from Industry 4.0 networks without significant performance degradation.

5. Practical Applicability: The framework will provide insights for deploying hybrid detection mechanisms in real-world industrial and enterprise settings.

8. Timeline

Phase	Duration
Literature Review	1 month
Data Collection & Preprocessing	1 month
Model Development	2 months
Model Training & Tuning	2 months
Performance Evaluation	1 month
Final Report Writing	1 month

9. Conclusion

In the era of Industry 4.0, where Cyber-Physical Production Systems (CPPSs) are integrated with smart technologies, the risk of cyber threats, particularly DDoS (Distributed Denial of Service) attacks, has become a significant concern. These attacks can severely disrupt the availability and performance of critical systems, leading to operational downtime and financial losses.

Traditional intrusion detection systems (IDS), relying on signature-based detection, are often ineffective against new and evolving attack patterns. This research addresses this gap by proposing a Hybrid Machine Learning (ML)-based model for DDoS attack detection in Industry 4.0 CPPSs. By combining supervised, unsupervised, and semi-supervised learning techniques, the hybrid model aims to enhance detection accuracy and minimize false positives in real-time environments.

The key contributions of this work include:

1. Improved detection accuracy for DDoS attacks, leveraging different learning models to capture both known and unknown attack behaviors.

2. A real-time detection system that is crucial for industrial environments, ensuring timely responses to threats.

3. An optimized balance between false positives and detection accuracy, ensuring minimal disruption to industrial processes.

The proposed model will be evaluated using public datasets like CICIDS 2019, and the results are expected to offer valuable insights into the effectiveness of hybrid ML models for securing Industry 4.0 systems. Further research could extend the work using real-world industrial data to assess the model's scalability and real-time performance in diverse industrial settings.

In conclusion, the Hybrid ML-based DDoS attack detection system provides an advanced and practical solution for securing Industry 4.0 CPPSs, ensuring better protection against cyber threats and supporting the resilience of critical industrial infrastructures.

References

1. Saghezchi, F.B., Mantas, G., Violas, M.A., Duarte, A.M., Rodriguez, J. ”Machine Learning for DDoS Attack Detection in Industry 4.0 CPPSs,” *Electronics 2022, 11, 602*. https://doi.org/10.3390/electronics11040602

2. CICIDS 2019 Dataset: Canadian Institute for Cybersecurity. Available at: https://www.unb.ca/cic/datase 2021.html

3. KDD Cup 99 Dataset: Available at: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

4. Zargar, S.T., Joshi, J., Tipper, D. ”A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks,” *IEEE Commun. Surv. Tutor.*, 2013.

5. Amouri, A., et al. ”Machine Learning for Intrusion Detection in IoT Networks,” *Sensors*, 2020.

6. Sarker, I.H., et al. ”Intrusion Detection using Machine Learning for IoT Networks,” *Sensors*, 2020.

7. Ribeiro, J., et al. ”An Autonomous Host-Based Intrusion Detection and Prevention System for Android Mobile Devices,” *IEEE Access*, 2020.

8. Zhang, Z., Zhang, X., Chen, L. ”A hybrid model for anomaly detection in industrial control systems,” *International Journal of Computer Applications*, 2019.

9. Linda, O., Vollmer, T., Manic, M. ”Neural Network Based Intrusion Detection System for Critical Infrastructures,” *2009 International Joint Conference on Neural Networks*.

10. Alhaidari, F.A., Al-Dahasi, E.M. ”A New Approach to Determine DDoS Attack Patterns on SCADA Systems,” *2019 ICCIS, Saudi Arabia*.

11. Lee, J., Bagheri, B., Kao, H.-A. ”A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems,” *Manufacturing Letters*, 2015.

12. Jafarian, M., Nasser, M., Ravanbakhsh, M., Kargar, M. ”Using machine learning to defend against DDoS attacks in IoT,” *Computers & Security*, 2020.

13. Pasqualetti, F., Dörfler, F., Bullo, F. ”Attack Detection and Identification in Cyber-Physical Systems,” *IEEE Trans. Autom. Control*, 2013.

14. Stouffer, K., Lightman, S., Pillitteri, V. ”Guide to Industrial Control Systems (ICS) Security,” *NIST*, 2015.

15. Perez, R.L., et al. ”Machine Learning for Reliable Network Attack Detection in SCADA Systems,” *IEEE Access*, 2018.