Project Progress Report
What advanced intrusion detection techniques can be implemented in SCADA systems to identify and mitigate unauthorized access or anomalies in real-time?

Savali Sandip Deshmukh
sdeshmukh@ucdavis.edu

Shivani Suryawanshi
ssuryawanshi@ucdavis.edu

Tanvi Mehta
tanmehta@ucdavis.edu

November 9, 2024

1 Introduction

Supervisory Control and Data Acquisition (SCADA) systems are essential for managing critical infrastructure like manufacturing, energy distribution, and water treatment. By enabling real-time monitoring, SCADA supports informed decision-making for operational efficiency. However, increased internet connectivity makes SCADA systems vulnerable to cyber threats, including buffer overflows, cross-site scripting, and SQL injection, which can disrupt public safety and operations.These attacks can compromise the availability, confidentiality, and integrity of critical data, leading to disastrous consequences in essential services. Among the most commonly targeted protocols in SCADA environments is the Modbus protocol, widely used for industrial communication but lacking intrinsic security features. This vulnerability makes Modbus an attractive target for unauthorized access and malicious activities.To address these risks, Intrusion Detection Systems (IDS) have become essential in detecting and mitigating threats in SCADA networks.

The primary goal of this research project is to:

- Analyze and evaluate advanced intrusion detection techniques applied to SCADA systems using the Gas Pipeline Dataset. By comparing we aim to identify the most effective methods for enhancing SCADA system resilience and security.
- Examine the common types of cyber-attacks on SCADA systems, particularly those exploiting vulnerabilities within the Modbus protocol.

2 Literature Review

The paper "Intrusion Detection and Identification System Design and Performance Evaluation for Industrial SCADA Networks" [1] focuses on the design and evaluation of an intrusion detection and identification system (IDIS) particularly for SCADA networks used in industrial settings. The authors recognize the critical vulnerability of SCADA systems to cyber-attacks due to their integration of information technology with operational technology, often resulting in significant operational and safety risks. To address this, they propose an IDIS framework that utilizes machine learning algorithms to detect and classify different types of intrusions effectively. Their methodology involves detailed analysis and feature selection to enhance detection accuracy, while the performance of the proposed system is benchmarked using real-world SCADA network data. The results demonstrate high detection rates, indicating that the system can effectively distinguish between normal operations and various intrusion scenarios. Additionally, the study provides insights into the computational efficiency of the system, highlighting its potential application in real-time SCADA environments. This research contributes significantly to enhancing cybersecurity in critical industrial infrastructure.

The paper "A Stacked Deep Learning Approach to Cyber-Attacks Detection in Industrial Systems: Application to Power System and Gas Pipeline Systems" [2] presents a comprehensive study on enhancing the detection of cyber-attacks in critical industrial systems through advanced machine learning techniques. The authors propose a novel, stacked deep learning framework that integrates multiple neural network models to improve the robustness and accuracy of intrusion detection. Focusing on power systems and gas pipeline infrastructures as case studies, the research highlights the growing need for effective cybersecurity measures in industrial environments vulnerable to complex cyber threats. The methodology involves leveraging deep learning's ability to capture non-linear relationships and subtle patterns in data to identify anomalies. The study's results show that the proposed approach outperforms traditional single-model systems, demonstrating superior performance in terms of both detection accuracy and false alarm reduction.

The paper "ICS-IDS: Application of Big Data Analysis in AI-Based Intrusion Detection Systems to Identify Cyberattacks in ICS Networks" [3] investigates the implementation of big data analytics in artificial intelligence-driven intrusion detection systems (IDS) for industrial control system (ICS) networks. The authors address the critical need for effective security mechanisms due to the increasing frequency and sophistication of cyber-attacks targeting ICS environments. Their proposed ICS-IDS framework leverages big data techniques to manage and analyze vast amounts of network data, ensuring timely and accurate threat detection. The study integrates AI models capable of distinguishing between normal and malicious activities, enhancing the system's capability to adapt to diverse attack patterns. Through extensive experimentation and performance evaluation, the results highlight the robustness and scalability of the framework in real-time intrusion detection.

3 Dataset

For this project, we plan to analyse the Gas Pipeline Dataset [4] for SCADA network security research created by Missouri University of Science and Technology (Missouri S&T). This dataset consists of 17 features and 274,628 instances across three class labels: binary, categorized, and specified. It includes 11 command payload features related to command injection attacks, 5 network features, and 1 response payload feature for response injection attacks.Some of the columns in the dataset are Timestamp, Source IP, Destination IP, Protocol, Packet Size, and Command Payload. The dataset covers 7 types of attacks which have been illustrated in figure below.

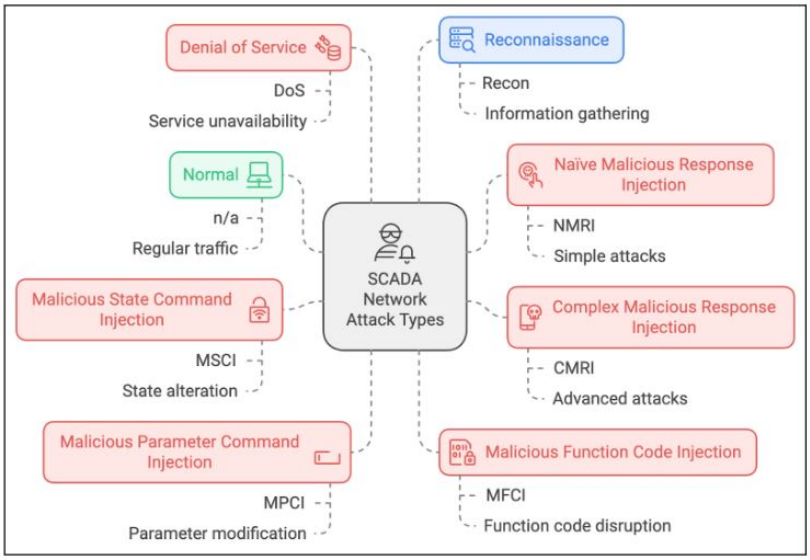

Figure 1: *Types of Attacks*

4 Methodology

The methodology for our Intrusion Detection System (IDS) on SCADA networks involves a structured workflow, as illustrated. We start with Data Collection and Preprocessing of the Gas Pipeline Dataset, which includes cleaning, handling missing values, and normalizing the data. After Data Splitting into training and testing sets, we proceed with Model Training using machine learning models such as Random Forest, Decision Trees, Support Vector Machines (SVM), and deep learning methods like Convolutional Neural Networks (CNNs), as used in previous research [1], [2], [3]. Next, we perform Hyperparameter Tuning to optimize model performance, followed by Model Testing on the test dataset. Finally, Performance Evaluation based on metrics like accuracy, precision, and recall to identify the most effective approach for intrusion detection on SCADA networks. This approach will enable a direct comparison of the different models' effectiveness.

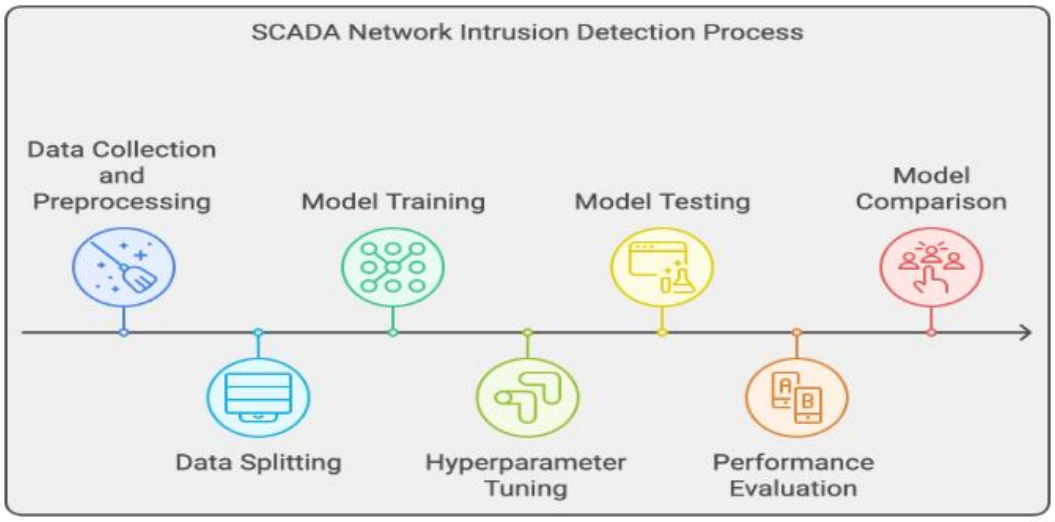

Figure 2: *Methodology*

5 Progress and Plans

So far, our research has progressed significantly. We have successfully identified a relevant dataset and several pertinent research papers that directly address our research question. By thoroughly examining the methodologies proposed in these papers, we have gained a deep understanding of the models and techniques employed. Additionally, we have delved into the theoretical underpinnings of SCADA systems, the potential vulnerabilities they face, and the specific reasons why the Modbus protocol is a primary target for cyberattacks. This foundational knowledge has allowed us to contextualize our research question and better comprehend the nature of the data we are working with.

Our initial efforts have focused on preprocessing the dataset to prepare it for analysis. Moving forward, we will partition the dataset into training and testing sets, enabling us to rigorously evaluate the performance of various machine learning models. Our objective is to implement the methodologies outlined in the research papers and strive to achieve comparable or superior accuracy rates.

6 Conclusion & Expected Results

In conclusion, the three papers provide valuable insights into the effectiveness of different intrusion detection models for Modbus-based industrial control systems. Khan et al. [1], uses machine learning-based anomaly detection, demonstrated that models such as Random Forest and Support Vector Machines (SVM) were effective in identifying unknown attacks, achieving accuracy rates of around 93%. Wang et al. [2], focused on misuse detection using a signature-based approach, with the proposed system detecting known attacks with high precision but lower recall rates, leading to an accuracy of about 87%. Ali et al. [3], implemented a hybrid model combining both anomaly detection and signature-based approaches, outperformed the others, achieving the highest accuracy of 95%. This indicates that combining machine learning techniques with traditional misuse detection methods yields the most robust results in detecting and mitigating Modbus cyberattacks, offering a more comprehensive solution for industrial control system security. We plan to implement and compare these models in a controlled experimental setting to validate these findings and identify the most suitable approach for our specific environment.

7 Appendix

Attack Type / Category / Class Name	Acronym	Brief Explanation of the names
Normal	n/a	Represents regular, non-malicious network traffic.
Naïve Malicious Response Injection	NMRI	Simple attacks that inject false responses into communication without complex behavior.
Complex Malicious Response Injection	CMRI	Advanced attacks that inject deceptive, sophisticated responses to disrupt communication.
Malicious State Command Injection	MSCI	Alters the state of devices by injecting unauthorized control commands.
Malicious Parameter Command Injection	MPCI	Injects malicious commands to modify parameters and operational settings of SCADA devices.
Malicious Function Code Injection	MFCI	Introduces unauthorized function codes to disrupt intended operations within SCADA protocols.
Denial of Service	DoS	Overloads the network or specific services to make them unavailable to legitimate users.
Reconnaissance	Recon	Probes the network to gather information on devices and configurations for potential future attacks.

Table 1: *Types of Attacks*

References

[1] Ahsan Al Zaki Khan Gursel Serpen. "Intrusion Detection and identification System Design and Performance Evaluation for Industrial SCADA Networks". In: (2020). DOI: https://doi.org/10.48550/arXiv.2012.09707.

[2] Wu Wang Fouzi Harrou. "A stacked deep learning approach to cyber-attacks detection in industrial systems: application to power system and gas pipeline systems". In: *Springer Nature* (2021). DOI: https://doi.org/10.1007/s10586-021-03426-w.

[3] Bakht Sher Ali Inam Ullah. "ICS-IDS: application of big data analysis in AI-based intrusion detection systems to identify cyberattacks in ICS networks". In: *Springer Nature* (2023). DOI: https://doi.org/10.1007/s11227-023-05764-5.

[4] Thomas Morris Wei Gao. "Industrial Control System Traffic Data Sets for Intrusion Detection Research". In: *Springer Nature* (2014). DOI: https://doi.org/10.1007/978-3-662-45355-1_5.