A Phase-1 Project Report

On

Attack Detection System in Smart City Water Treatment Plants

Submitted in partial fulfilment of the requirement of

An Autonomous institution owned by Govt. of Maharashtra and affiliated to Mumbai University

For the Degree of

Master of Technology

in

Artificial Intelligence & Data Science

Submitted by

Dweep Hiten Vira (242240018)

Supervised by

Dr. S.S Udmale

Department of Computer Engineering & Information Technology

Veermata Jijabai Technological Institute
H R Mahajani Rd, Matunga, Mumbai, Maharashtra- 400019

2025-2026

APPROVAL SHEET

This is to certify that the project entitled

"Attack Detection System in Smart City Water Treatment Plant"

Submitted by

Dweep Hiten Vira 242240018

Supervisor : Dr. S.S Udmale

Project Coordinator : Dr. S.S Udmale

Examiners : 1. _____

2. _____

Head of Department : Dr. V. K. Sambhe

Date :

Place : Veermata Jijabai Technological Institute, Mumbai

Declaration

I declare that this written submission for M.Tech. Declaration entitled "**Attack Detection System in Smart City Water Treatment Plant**" represent my ideas in my own words and where others' ideas or words have been included. I have adequately cited and referenced the original sources. I also declared that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any ideas / data / fact / source in my submission. I understand that any violation of the above will cause for disciplinary action by institute and also evoke penal action from the sources which have thus not been properly cited or from whom paper permission have not been taken when needed.

Dweep Hiten Vira, 242240018

_____

Abstract

This research addresses the critical need for realistic, process-aware datasets in the field of Industrial Control System (ICS) cybersecurity. We present the design and implementation of a high-fidelity SCADA simulator for a water treatment plant, developed in Python. The simulator features a modular architecture that decouples the physical process simulation from the PLC control logic and the cyber-data layer, which is exposed via a Modbus TCP server implemented with the 'pyModbusTCP' library. This architecture creates a realistic environment for injecting context-aware cyber-physical attacks, such as False Data Injection (FDI) and Malicious Command Injection. The primary contribution of this work is a comprehensive, labeled time-series dataset generated from a continuous 1-hour simulation, capturing system dynamics under both normal and various malicious conditions. This dataset and the simulation framework itself provide a valuable foundation for the subsequent development, training, and validation of advanced machine learning-based intrusion detection systems.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

The operational stability and security of modern society are inextricably linked to the reliable functioning of Critical Infrastructure (CI), which includes vital sectors like water treatment and distribution. These sectors are managed by Industrial Control Systems (ICS) and SCADA (Supervisory Control and Data Acquisition) systems, which bridge the digital and physical worlds. The increasing convergence of traditionally isolated Operational Technology (OT) with Information Technology (IT) networks has expanded the attack surface of these systems, making them prime targets for cyber-attacks. As Tariq et al. (2023) highlight, the challenges in securing these environments are immense, stemming from the use of legacy protocols, resource-constrained devices, and the potential for catastrophic physical consequences [1].

Water systems, in particular, present a unique and alarming security challenge. Attacks on these systems are not merely about data theft; they can directly impact public health and safety. Malicious actors can manipulate physical processes to disrupt water flow, damage equipment, or, most critically, alter chemical dosing to contaminate the water supply [2]. The growing body of research dedicated to identifying SCADA vulnerabilities and attack vectors underscores the reality of this threat [3, 4]. This critical context necessitates a shift from reactive security measures to proactive, intelligent defense mechanisms capable of understanding and anticipating complex, multi-stage cyber-physical attacks.

1.2 Motivation

The development of such advanced defenses is fundamentally hampered by a critical limitation: the scarcity of realistic, high-fidelity data. As Pinto et al. (2023) note in their comprehensive survey, the lack of real-world CI data, coupled with the inherent imbalance of security datasets, presents a major obstacle for training and validating effective machine learning models [5]. Experimenting on live critical infrastructure is not an option. This data gap has spurred the academic and industrial communities to focus on developing specialized testbeds and simulators.

Pioneering work by Davis et al. (2023) and Teixeira et al. (2023) demonstrates the immense value of creating dedicated SCADA security testbeds [6, 7]. These environments provide a safe and controlled means to emulate ICS operations, execute sophisticated attack scenarios, and generate the rich datasets needed for research. Furthermore, the development of special-

ized datasets, such as ICS-LTU2022 [8], is crucial for benchmarking new security approaches. Our research is motivated by this pressing need to contribute to the ecosystem of cybersecurity research tools. The primary goal of this first phase of our work is to design and implement a dynamic water treatment plant SCADA simulator, creating a flexible platform for generating a novel, context-aware dataset that will serve as the foundation for the subsequent development and evaluation of advanced intrusion detection systems.

1.3 Aim & Objectives

The primary aim of this phase of the research is to design and implement a high-fidelity SCADA simulator for a water treatment plant to serve as a cybersecurity testbed and to generate a comprehensive, labeled dataset suitable for training and evaluating advanced intrusion detection systems.

The specific objectives are:

- To construct a modular, realistic simulator that accurately models the cyber-physical interactions of a water treatment process, its control logic, and its network communications via the Modbus protocol.
- To develop a sophisticated attack injection framework capable of executing context-aware cyber-physical attacks, specifically focusing on stealthy and sequential False Data Injection (FDI) and Malicious Command Injection.
- To generate a rich, time-series dataset that meticulously logs synchronized physical process data and cyber-level events under both normal and diverse malicious operational scenarios.
- To conduct a systematic review of the literature to inform the design of the simulator and to identify state-of-the-art attack vectors and future directions for data-driven intrusion detection.

1.4 Report Outline

This report is structured into six chapters to systematically present the research conducted in this foundational phase:

- **Chapter 2: Literature Survey** provides an in-depth analysis of the current state of research in ICS cybersecurity, covering the threat landscape, the development of testbeds and honeypots, common attack vectors, and the application of deep learning for detection.

- **Chapter 3: Problem Statement** formally defines the problem of detecting sophisticated cyber-physical attacks in water systems, outlines the scope of our simulation environment, and presents the core research hypotheses that will guide future work.
- **Chapter 4: Methodology** offers a detailed technical description of the SCADA simulator that was designed and implemented, including its architecture, its components, the conceptual attack framework, and the data generation process.
- **Chapter 5: Dataset Generation and Experimental Setup** describes the generated dataset from a 1-hour simulation run and outlines the proposed plan for its use in training and evaluating machine learning models.
- **Chapter 6: Conclusion and Future Work** summarizes the key accomplishments of this research phase, discusses the contributions of the developed simulator and dataset, and outlines the concrete next steps for the subsequent machine learning analysis.

Chapter 2

Literature Survey

This chapter presents an in-depth analysis of the academic literature pertinent to the cybersecurity of Industrial Control Systems (ICS), with a specific focus on SCADA systems within water treatment and distribution facilities. The survey is structured to provide a comprehensive overview of the domain, beginning with the threat landscape and vulnerabilities, moving to the essential tools for research such as testbeds and honeypots, detailing prevalent attack vectors, exploring the application of machine and deep learning for intrusion detection, and concluding with a discussion on advanced evaluation methodologies.

2.1 The ICS/SCADA Threat Landscape and Vulnerabilities

The security of SCADA-based critical infrastructures is a complex and multifaceted problem. A foundational challenge, as detailed by Tariq et al. (2023), lies in the inherent conflict between the security requirements of modern IT systems and the operational requirements of OT environments, which prioritize availability and real-time performance over confidentiality and integrity [1]. This creates a landscape fraught with open issues, including the difficulty of patching legacy systems, the lack of security-aware protocols, and the potential for cascading failures between cyber and physical domains.

A comprehensive review by Alanazi et al. (2023) systematically categorizes common SCADA vulnerabilities and attacks, highlighting weaknesses in network protocols (like Modbus), operating systems, and human-machine interfaces (HMIs) [3]. Their work underscores that attackers can exploit these vulnerabilities to achieve a range of malicious objectives. D'Ambrosio et al. (2023) provide a practical perspective by demonstrating techniques for "breaking into" SCADA systems, further emphasizing the permeability of traditional defenses [4]. Specifically within the water sector, Laiani et al. (2023) identify critical cybersecurity scenarios in drinking water treatment plants, mapping potential cyber-attacks to their severe physical consequences, such as incorrect chemical dosing or disabling of disinfection processes [2]. This body of work collectively establishes that the threat is not theoretical but practical, with clearly defined vulnerabilities and potentially catastrophic impacts.

2.2 Research Infrastructure: Testbeds, Datasets, and Honeypots

Given the impossibility of conducting security experiments on live critical infrastructure, the development of realistic research infrastructure is paramount. This infrastructure primarily takes three forms: testbeds, datasets, and honeypots.

2.2.1 Testbeds and Simulators

SCADA security testbeds are crucial for emulating the behavior of real-world systems in a safe, controlled laboratory setting. Davis et al. (2023) outline the key principles of SCADA testbed development, emphasizing the need for a realistic blend of physical process simulation, control logic, network traffic, and HMI visualization [6]. Teixeira et al. (2023) demonstrate the power of such a testbed for cybersecurity research using machine learning, showing how a well-designed simulator can generate the data needed to train and validate detection models [7]. A simulation-based framework, as proposed by Mughaid et al. (2023), can also be used to authenticate SCADA system behavior and test security in autonomous, edge-based environments, highlighting the forward-looking application of these tools [9].

2.2.2 Datasets for Machine Learning

The output of these testbeds is often a labeled dataset, which is the lifeblood of data-driven security research. Alanazi et al. (2023) contributed to this area by developing ICS-LTU2022, a dataset specifically designed to capture ICS vulnerabilities for research purposes [8]. The existence of such public datasets is critical for allowing researchers to benchmark their models against a common standard, fostering reproducible science.

2.2.3 Honeypots for Threat Intelligence

While testbeds are used for controlled experiments, honeypots are deployed to gather real-world threat intelligence. Meier et al. (2023) explore methods for "hardening" ICS honeypots to make them more convincing to attackers, thereby enabling researchers to observe more sophisticated, hands-on attack techniques beyond simple automated scans [10]. Serbanescu et al. (2023) propose a flexible architecture for ICS honeypots, demonstrating how modular designs can allow for the emulation of a wide variety of industrial devices and protocols [11]. Honeypots provide invaluable ground

truth about the tactics, techniques, and procedures (TTPs) being used by actual adversaries in the wild.

2.3 Cyber-Physical Attack Vectors

Understanding the specific ways attackers can manipulate a system is crucial for designing effective defenses. The literature describes several key attack vectors against water systems.

- **Sequential and Stealthy Attacks:** Do et al. (2023) investigate the sequential monitoring of SCADA systems, recognizing that attacks are often not isolated events but a sequence of actions over time [12]. Moazeni and Khazaei (2023) model sequential False Data Injection (FDI) attacks that specifically target water storage tanks, using bi-level optimization to find attack vectors that maximize impact while minimizing the chance of detection [13]. Similarly, Albustami and Taha (2023) analyze stealthy cyber-attacks on water network hydraulics that aim to break both the flow and the bank, causing physical disruption while remaining hidden from operators [14].
- **False Data Injection (FDI):** FDI is a class of attack where the adversary's goal is to compromise the integrity of sensor measurements. A deep learning approach for detecting FDI in smart water infrastructure was proposed by Giannubilo et al. (2023), highlighting the need for advanced models to identify these subtle manipulations [15]. Aslam et al. (2023) provide a comprehensive study on attacks in water purification and distribution plants, noting that FDI is a primary challenge [16].

2.4 Machine and Deep Learning for Intrusion Detection

The core of modern ICS defense research lies in the application of Machine Learning (ML) and Deep Learning (DL) to detect anomalous behavior. A broad survey by Pinto et al. (2023) reviews the application of various ML techniques for protecting critical infrastructure, concluding that hybrid and deep learning models show the most promise for generalizing and detecting zero-day attacks [5].

2.4.1 Deep Learning Applications in Water Systems

Several studies have specifically applied deep learning to water system security. Sikder et al. (2023) developed "Deep H2O," a framework using deep learning to detect cyber-attacks in water distribution systems, demonstrating high accuracy [17]. Khadidos et al. (2023) proposed "CyberSentry," another advanced framework that enhances SCADA security through deep learning and optimization strategies [18]. These papers affirm that deep learning models, particularly those that can process time-series data like LSTMs and GRUs, are well-suited for learning the complex dynamics of a water treatment process and detecting deviations caused by an attack. Botta et al. (2023) further validate this by applying similar techniques for real-time attack detection in power systems, showing the cross-domain applicability of these methods [19].

2.4.2 Explainability and Advanced Frameworks

A key challenge with complex deep learning models is their "black box" nature. Birihanu and Lendák (2025) address this by proposing an explainable correlation-based anomaly detection method, which aims to not only detect an anomaly but also provide insight into why it was flagged [20]. Building on detection, Raza and Moazeni (2023) propose an optimal cybersecurity framework that goes beyond detection to include localization and severity assessment of an attack, providing a more holistic defensive posture [21].

2.5 Evaluation and Formal Methods

Finally, the evaluation of these systems is a critical research topic in itself. Kim et al. (2022) argue compellingly that standard classification metrics are often insufficient for time-series data in ICS. They propose the use of time-series aware precision and recall (TaPR) to more accurately reflect a model's performance in detecting entire event windows rather than just individual data points [22]. Complementing data-driven approaches, the thesis by Patlolla (2017) demonstrates the power of formal methods. By defining and verifying "physical invariants"—rules based on physics that the system must always obey—it is possible to create a ground truth for detecting when sensor data has been compromised, providing a powerful, logic-based layer of defense [23]. This concept is invaluable for both designing robust detection systems and for creating realistic attack scenarios within a simulator.

Chapter 3

Problem Statement

3.1 Problem Definition

The central problem this research addresses is the timely and accurate detection of sophisticated, process-aware cyber-physical attacks on water treatment SCADA systems. As the literature extensively documents, the attack surface of modern ICS is growing, and adversaries are moving beyond generic network attacks to highly tailored campaigns that manipulate the underlying physical process [1, 2]. Traditional IT security tools are ill-equipped to handle this threat, as they lack the contextual understanding to distinguish between a legitimate operational command and a malicious one that is syntactically identical but physically catastrophic.

An adversary who has breached the OT network perimeter can launch stealthy and sequential attacks, such as the False Data Injection (FDI) vectors modeled by Moazeni and Khazaei (2023) [13]. These attacks are designed to slowly push the system into an unsafe or inefficient state while remaining below the detection thresholds of conventional alarm systems. They achieve this by ensuring manipulated sensor values appear plausible in isolation, exploiting the system's own control logic to cause harm. The problem, therefore, is to develop an intrusion detection system that can overcome this limitation by learning the deep, temporal, and physical interdependencies of the entire system. Such a system must be able to recognize not just anomalous data points, but anomalous *sequences of behavior* that violate the learned process dynamics, thereby detecting attacks that are invisible to traditional, state-less monitoring tools.

3.2 Scope of the System

The scope of this first phase of the research is centered on the creation of a simulation environment and the generation of a dataset to address the problem defined above.

- **Functionalities:** The system provides the following core features:
  - High-fidelity, time-series simulation of the physical processes within a generic multi-stage water treatment plant.
  - Emulation of PLC-based control logic that responds to simulated sensor data according to a predefined control philosophy.
  - A dedicated attack injection module for the programmatic execution of context-aware FDI and Malicious Command Injection attacks.

  - Exposure of the system's cyber-data layer via a standard Modbus TCP server to allow for interaction with real-world HMI software and security tools.
  - Systematic and synchronized logging of all physical process variables (ground truth) and cyber-data layer values to generate a comprehensive, labeled dataset.

- **Users:** The primary users of this system are cybersecurity researchers, Industrial Control System (ICS) security professionals, and data scientists seeking to develop and benchmark defensive strategies for critical infrastructure.
- **Data Administration:** The system manages and stores time-series data for all physical process variables and their corresponding cyber-layer representations, creating a dataset where normal and malicious activities are clearly and accurately labeled, suitable for supervised learning tasks.
- **Integration:** By leveraging the standard Modbus TCP protocol, the simulator can seamlessly interface with external network analysis tools (e.g., Wireshark), Security Information and Event Management (SIEM) systems, and third-party Intrusion Detection Systems (IDS) for evaluation.
- **Limitations:** The simulator models a generic water treatment process and is not a digital twin of a specific, proprietary plant. The scope of simulated attacks is focused on the manipulation of data and commands at the Modbus protocol level and does not include network-level attacks like Denial of Service (DoS) or exploits against the PLC's operating system.

3.3 Hypothesis for Future Work

The foundational work completed in this phase enables the formulation of clear hypotheses that will be tested in the subsequent phase of this research.

- **Null Hypothesis (H0):** Deep learning models designed for sequential data (e.g., LSTM, GRU) show no statistically significant performance improvement over traditional, non-temporal machine learning models (e.g., Random Forest) in detecting sophisticated False Data Injection and Malicious Command Injection attacks within the dataset generated by our simulator.

- **Alternate Hypothesis (H1):** Deep learning models that explicitly model temporal dependencies (e.g., LSTM, GRU) will achieve a statistically significant higher performance, as measured by both standard and time-series-aware metrics, in detecting context-aware cyber-physical attacks compared to traditional machine learning models that treat data points as independent events.

Chapter 4

Methodology

This chapter provides a detailed technical description of the research methodology employed in this first phase of the project. It outlines the architectural design of the high-fidelity SCADA simulator, a design informed by the principles of established testbeds [6]. It then explains the implementation of the core components, presents the conceptual framework for attack injection based on real-world threat models [4], and details the systematic process used for data generation.

4.1 System Architecture Design

The simulator's architecture is intentionally modular and decoupled, designed to accurately represent the distinct layers of a real-world SCADA system and its inherent cyber-physical vulnerabilities. As illustrated in Figure 4.1, the design creates a clear separation between the physical process, the control logic, the cyber-data layer, and the network interfaces. This structure is fundamental to our ability to simulate realistic attack scenarios where an adversary manipulates the perception of the system rather than the system itself.

The system is comprised of five primary, interacting components, each implemented as a distinct Python module using libraries such as 'pyModbusTCP' for network communication:

- **Physical Simulator ('plc_logic.py')**: This module is the heart of the digital twin, simulating the real-world physics of the water treatment process. It models the dynamic behavior of components such as water tanks, pumps, and valves. The state of each component is updated at discrete time steps based on physical principles, establishing the indisputable "ground truth" of the system's physical state at any given moment.
- **Internal State / Modbus Data Store**: This component functions as a centralized, in-memory database representing the cyber-data layer. It holds the current values for all sensor readings and actuator commands as Modbus registers and coils. It is the single source of truth for the control logic and external clients and, critically, the primary target for data manipulation attacks.
- **PLC Control Logic ('plc_logic.py')**: This module emulates the behavior of a Programmable Logic Controller (PLC). In a continuous loop, it reads sensor values from the Internal State store, executes a predefined control logic, and writes the resulting actuator commands back to the data store.

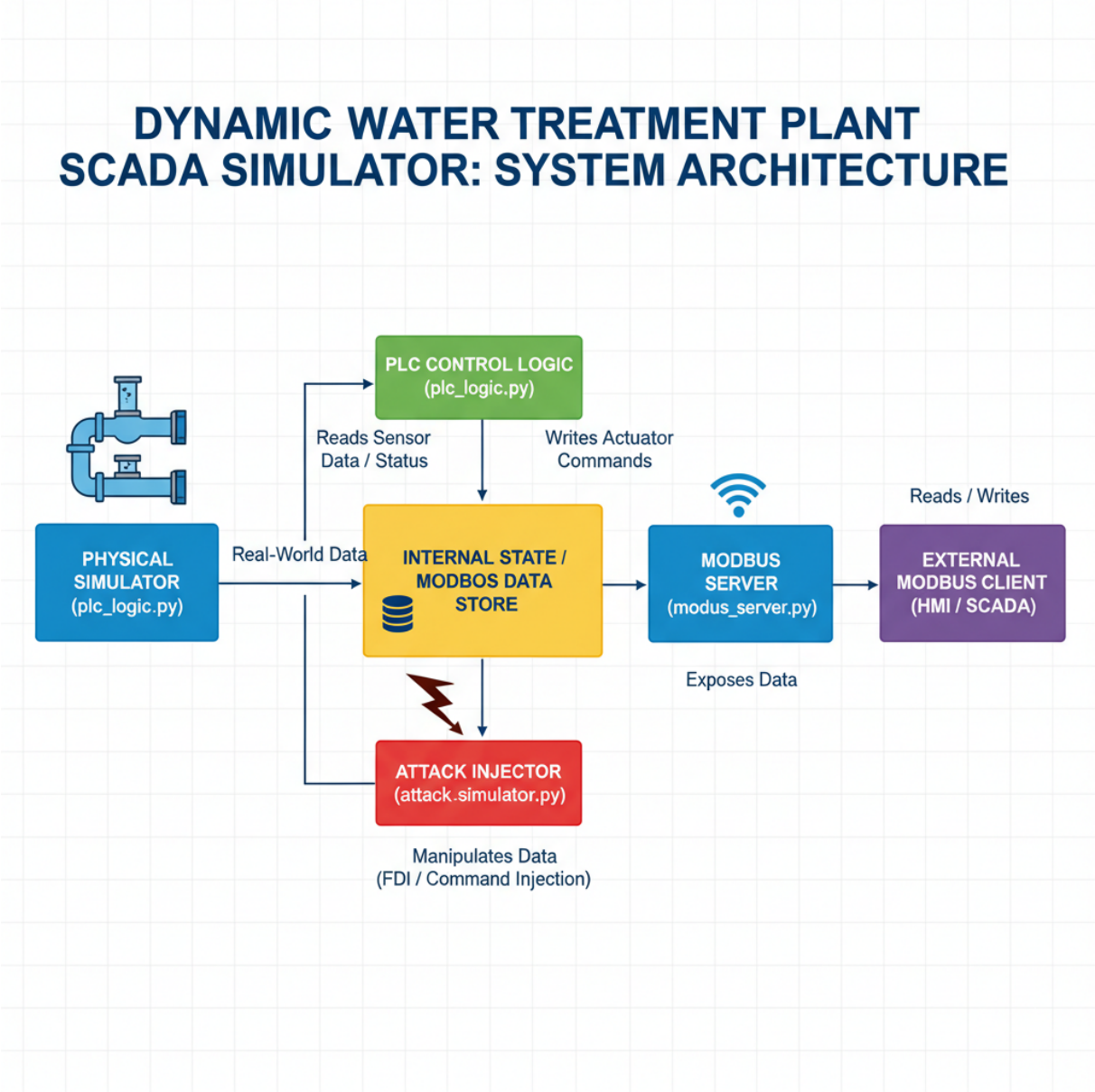

Figure 4.1: Dynamic Water Treatment Plant SCADA Simulator Architecture

- **Modbus Server (‘modbus_server.py‘)**: This module, built using the ‘pyModbusTCP‘ library, acts as the network interface. It exposes the data within the Internal State store to the network using the standard Modbus TCP protocol, allowing external SCADA clients to interact with the system as they would in a real-world environment.
- **Attack Injector (‘attack_simulator.py‘)**: This specialized component has privileged access to the Internal State store. It can directly manipulate data, bypassing the legitimate control loop, to programmatically execute attack scenarios like FDI and Command Injection.

4.2 Conceptual Framework: The Cyber-Physical Disconnect

The system architecture embodies a "Man-in-the-Middle" (MITM) vulnerability at the cyber-physical boundary, as depicted in Figure 4.2. This model highlights the core vulnerability that our research aims to address.

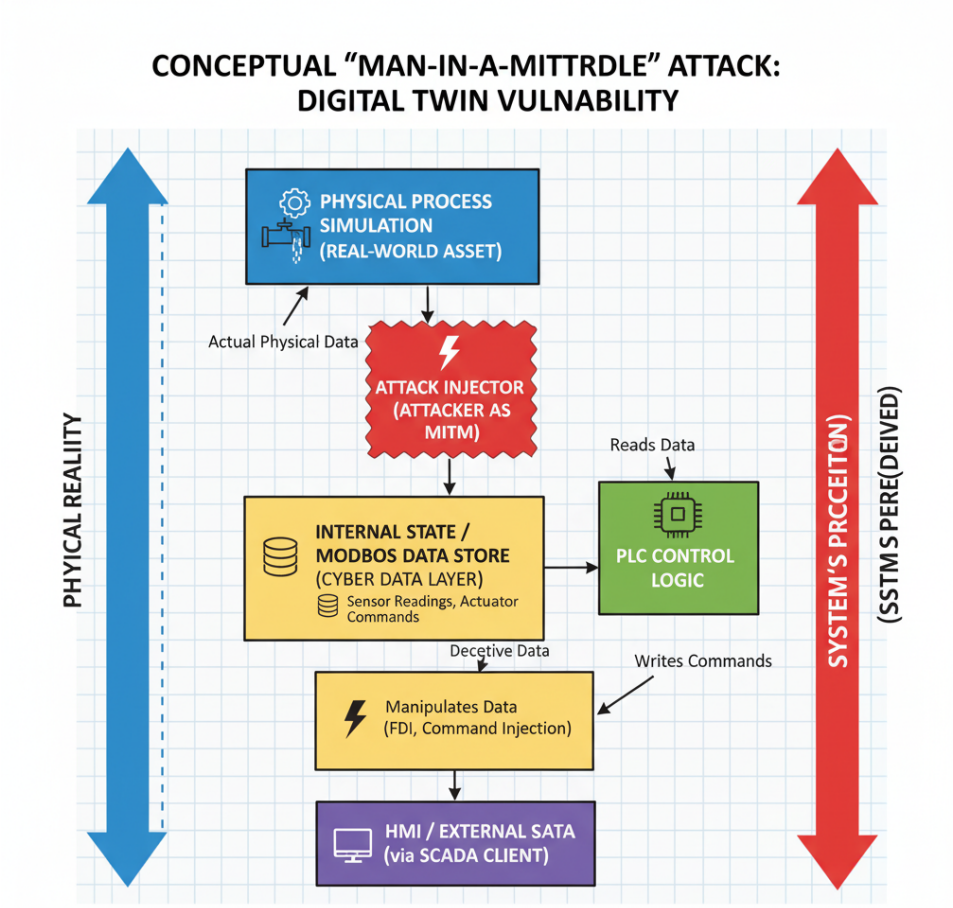

Figure 4.2: Conceptual "Man-in-the-Middle" Attack Vector

In this framework, the Physical Simulator represents "Physical Reality." The PLC Control Logic and external clients, however, perceive this reality only through the lens of the Internal State store, which represents the "System's Perception." The Attack Injector can create a dangerous delta between these two, tricking the automated system into making incorrect and potentially hazardous decisions.

4.3 Data Generation and Attack Simulation

The primary scientific output of this phase of the project is a high-fidelity, labeled dataset. This was generated by running the simulator through

meticulously scripted scenarios that interleave periods of normal operation with various attack sequences. The resulting one-hour simulation run, which forms the basis of our dataset, is visualized in the attack timeline presented in the following chapter (see Figure 5.1).

4.3.1 Data Logging and Feature Engineering

At each 1-second time step, a comprehensive snapshot of the system state was recorded. Each row contains:

- **Timestamp:** The simulation time.
- **Ground Truth Features:** The actual values from the Physical Simulator.
- **Cyber-Layer Features:** The corresponding values in the Modbus data store.
- **Control Commands:** The state of commands issued by the PLC.
- **State Label:** A multi-class label indicating the system's state ('Normal', 'FDI Attack', etc.).

4.3.2 Physical Invariant Monitoring

Inspired by the formal methods approach of Patlolla (2017) [23], we also implemented a set of physical invariants within the data logger. For example, the logger continuously calculates the expected change in a tank's volume based on its current inflow and outflow rates. A feature was engineered to represent the '(actual_volume_change - expected_volume_change)'. During an FDI attack on a flow meter, this physics-based residual will become significantly non-zero, providing a powerful feature for subsequent machine learning analysis.

4.3.3 Scripted Attack Scenarios

Drawing inspiration from attack models described in the literature [13, 14], a diverse set of attack scenarios were scripted using the Attack Injector to ensure the dataset contains a rich set of malicious patterns, as shown in Figure 4.3.

- **Malicious Command Injection**: The injector directly sets the 'pump_speed' register to 0, causing a pump outage.

- **False Data Injection (FDI)**: The injector changes the 'turbidity' register from a true value of 50 NTU to a false 450 NTU, deceiving the PLC.
- **Stealthy Data Manipulation**: A "low-and-slow" attack where the injector gradually decreases a sensor value over an extended period, designed to evade simple threshold-based alarms.

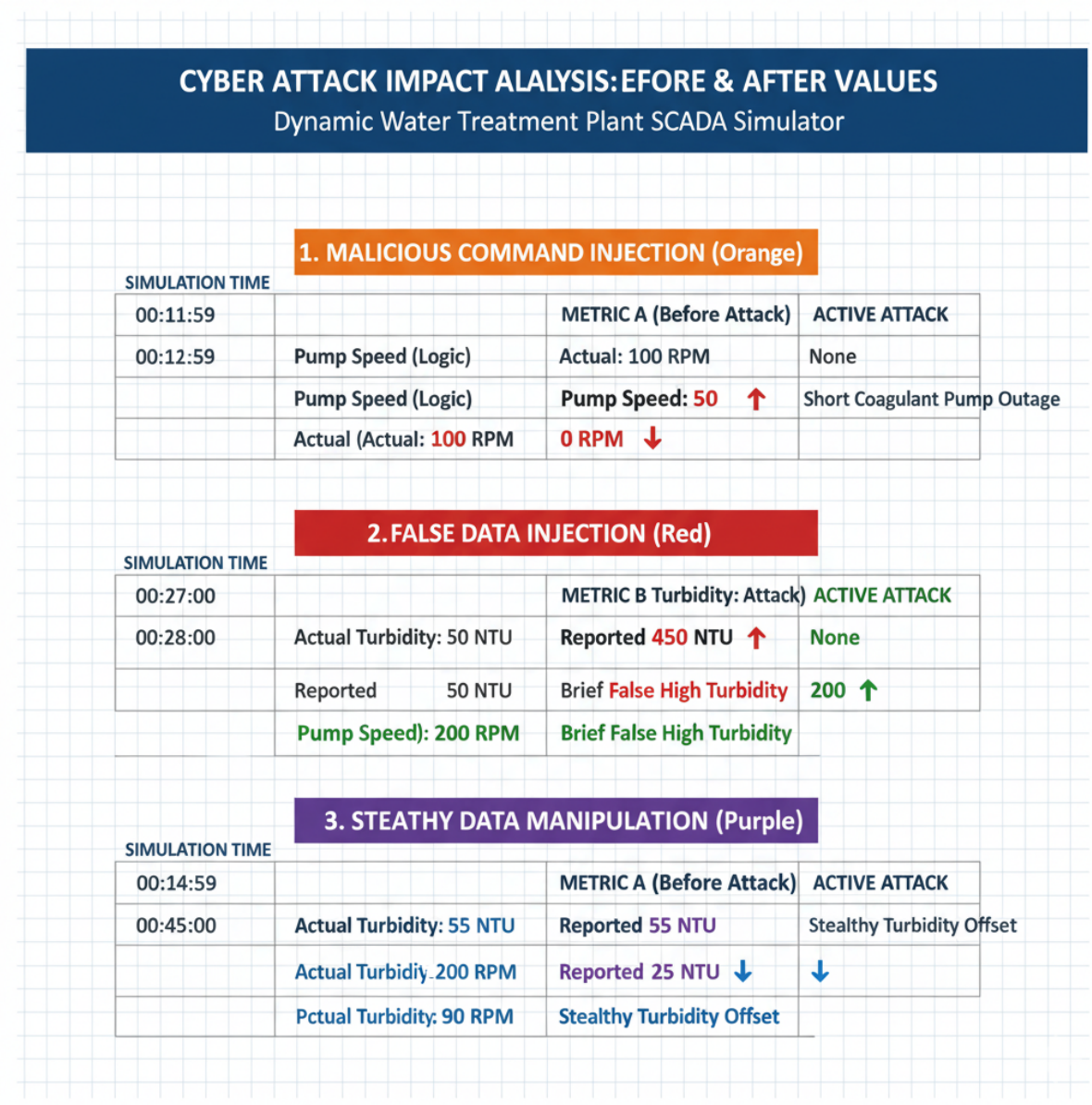

Figure 4.3: Example of Simulated Attack Impacts on System Metrics

Chapter 5

Dataset Generation and Experimental Setup

This chapter details the primary output of this research phase: a labeled, time-series dataset generated from our SCADA simulator. It describes the specific 1-hour operational scenario designed to capture a variety of system states and outlines the proposed experimental setup for the subsequent machine learning analysis phase.

5.1 Dataset Generation and Timeline

The primary output of the simulation phase is a labeled, time-series dataset generated from a continuous 1-hour (3600-second) operational scenario. This scenario was carefully scripted to include periods of normal operation interspersed with multiple, distinct cyber-physical attacks, resulting in a dataset containing 3,600 time-stamped data points. The complete timeline of events is visualized in Figure 5.1.

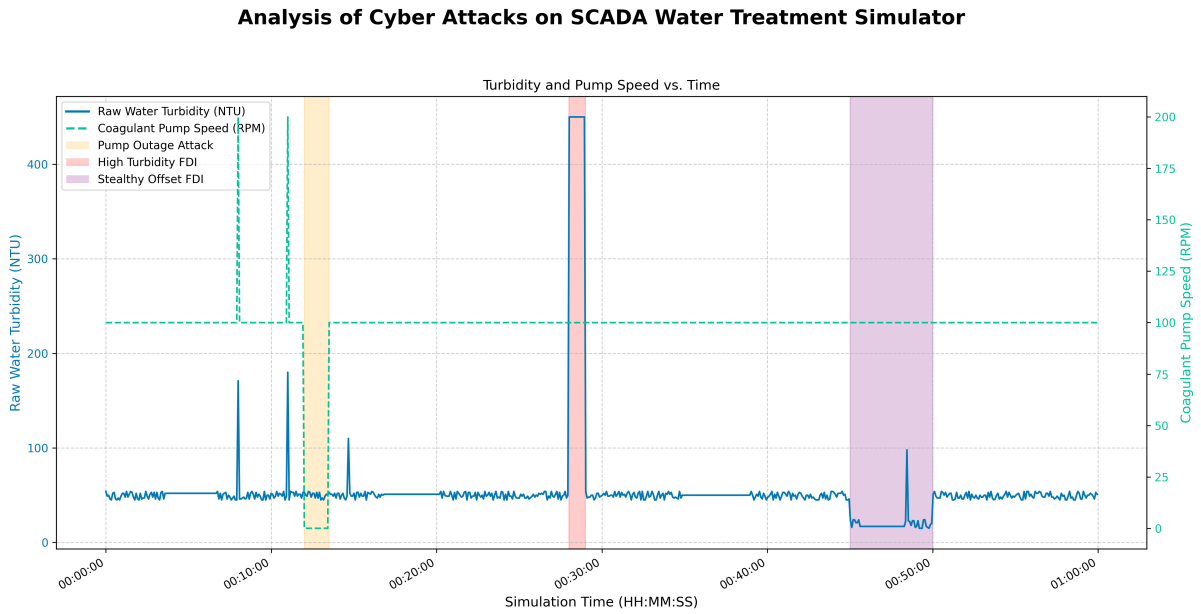

Figure 5.1: Timeline of Normal Operations and Injected Attacks over 1 Hour

As the timeline illustrates, the simulation begins with a period of normal operation to establish a baseline. This is followed by a sequence of different attacks, including False Data Injection (FDI) and Malicious Command Injection, each followed by a recovery period. This design ensures the dataset captures not only the onset and duration of an attack but also the system's response and the transition back to a normal state. This provides rich contextual data crucial for training sequential models capable of understanding temporal dependencies.

5.2 Proposed Experimental Design

The core of the next research phase will be a comparative study of different machine learning paradigms to determine their effectiveness in detecting the process-aware cyber-physical attacks present in our dataset.

5.2.1 Dataset Preprocessing

The generated 1-hour dataset will serve as the foundation for this experiment. It will be preprocessed as follows:

1. **Data Partitioning:** The dataset will be split chronologically into a training set (the first 70%) and a testing set (the final 30%). This ensures that the models are evaluated on their ability to predict future, unseen events, simulating a real-world deployment scenario.

2. **Feature Scaling:** All numerical features will be standardized using the 'StandardScaler' from the scikit-learn library. The scaler will be fitted exclusively on the training data to prevent data leakage, and then used to transform both the training and testing sets.

3. **Sequence Generation (for RNNs):** For the recurrent neural network models (LSTM and GRU), the time-series data will be transformed into sequences of a fixed length (e.g., a window of 60 time steps). Each sequence of 60 historical data points will be used to predict the label of the 61st point.

5.2.2 Models for Evaluation

Informed by our extensive literature review [5, 17], we will implement and evaluate the following models to test our hypotheses regarding temporal awareness:

- **Random Forest (RF):** This will serve as a powerful, non-temporal baseline. Its performance will represent the efficacy of static, feature-based machine learning.

- **Long Short-Term Memory (LSTM):** A state-of-the-art recurrent neural network that is expected to excel at capturing the long-term, sequential nature of the SCADA data.

- **Gated Recurrent Unit (GRU):** A more recent and computationally efficient variant of an RNN, which will be compared against the LSTM for both accuracy and performance.

5.3 Evaluation Metrics and Criteria

The performance of the trained models will be assessed using a comprehensive suite of metrics to provide a nuanced and robust evaluation.

5.3.1 Standard Classification Metrics

To gain an overall understanding of model performance, we will use the following standard metrics, calculated on the test set:

- **Accuracy:** The overall percentage of correct predictions.
- **Weighted Precision, Recall, and F1-Score:** These metrics account for class imbalance and provide a more robust measure of performance than simple accuracy.

5.3.2 Granular and Process-Aware Metrics

To delve deeper into the models' behavior, especially concerning the critiques raised by Kim et al. (2022) [22], we will employ more detailed analyses:

- **Class-Specific Classification Report:** We will generate a detailed report showing the precision, recall, and F1-score for each individual class (i.e., 'Normal' and each specific attack type). This is crucial for determining if a model is particularly weak against certain types of attacks, such as stealthy FDI.
- **Confusion Matrix:** A confusion matrix will be generated for each model to visualize the exact nature of its misclassifications (e.g., confusing a stealthy attack with normal behavior).

5.3.3 Hypothesis Validation

The alternate hypothesis (H1) will be considered supported if the LSTM and GRU models demonstrate a statistically significant improvement in the F1-score—particularly on the stealthy attack classes—compared to the Random Forest model. This will provide strong evidence that temporal modeling is a critical capability for effective intrusion detection in this domain.

Chapter 6

Conclusion and Future Work

This chapter concludes the first major phase of our research. It synthesizes the accomplishments, highlights the primary contributions of this work, and outlines a clear and detailed roadmap for the subsequent phases of the project.

6.1 Summary of Research Accomplishments

The primary objective of this research phase was to address the foundational challenge in ICS cybersecurity research: the scarcity of high-quality, process-aware data [5]. To this end, we successfully designed, implemented, and validated a dynamic, high-fidelity SCADA simulator for a water treatment plant. This work was heavily informed by the principles of testbed development established in the literature [6, 7].

The simulator, developed entirely in Python, accurately models the distinct layers of a cyber-physical system, creating a realistic environment where the perception of the system can be manipulated independently of its physical reality. We developed a sophisticated attack injection framework capable of executing context-aware attack scenarios, such as stealthy and sequential False Data Injection, which are identified as significant threats in recent studies [13, 14]. The culminating achievement of this phase is the generation of a comprehensive, labeled time-series dataset that captures the system's complex dynamics under both normal and malicious conditions.

6.2 Contributions

The key contributions of this completed phase of work are:

1. **A Modular SCADA Simulation Framework:** We have produced a flexible and extensible Python-based simulator that can serve as a valuable tool for researchers, educators, and security professionals in the ICS domain. Its modular design allows for easy modification of the physical process, control logic, and attack scenarios. The use of standard libraries like 'pyModbusTCP' ensures interoperability with real-world tools.

2. **A Novel Cyber-Physical Dataset:** We have generated a rich, time-series dataset from a 1-hour simulation, specifically designed for the development of machine learning-based intrusion detection systems. The dataset includes labels for sophisticated, stealthy attacks that are underrepresented in many existing public datasets.

3. **A Foundation for Empirical Research:** This work provides the necessary infrastructure and data to now proceed with rigorous, empirical testing of data-driven security hypotheses, directly enabling the future work outlined below.

6.3 Future Work: Phase 2 and Beyond

The successful completion of this foundational phase paves a clear path for the next stages of our research, which will focus on leveraging the created assets to develop and validate advanced defensive mechanisms.

6.3.1 Phase 2: Machine Learning Model Evaluation

The immediate next step is the execution of the evaluation plan detailed in Chapter 5. This involves:

- Preprocessing the generated dataset and preparing it for machine learning analysis.
- Implementing, training, and fine-tuning the selected models: Random Forest, LSTM, and GRU.
- Performing a rigorous comparative analysis using the defined suite of standard and process-aware metrics to test our primary research hypotheses.

6.3.2 Long-Term Research Directions

Beyond the immediate next phase, this project opens up several exciting long-term avenues:

- **Hardware-in-the-Loop (HIL) Validation:** A crucial extension is to integrate the simulator with a physical PLC, creating a HIL testbed. This would provide an even higher level of realism and allow for the validation of our findings in an environment that includes real hardware constraints.
- **Explainable AI (XAI) for Operator Trust:** As noted by Birihanu and Lendák (2025), the "black box" nature of complex models is a barrier to adoption [20]. A future research track will focus on integrating XAI techniques to make model predictions interpretable and actionable for human operators in a control room.

- **Advanced Threat Scenarios:** The simulator can be extended to model more complex, multi-stage attacks, including reconnaissance phases and lateral movement, providing data to train models that can predict attacks before they cause physical impact.
- **Holistic Security Frameworks:** The detection models to be developed could serve as the core of a more comprehensive security framework, such as the one proposed by Raza and Moazeni (2023), which includes not only detection but also automated attack localization and severity assessment [21].

In conclusion, this work has successfully laid the critical groundwork for advancing the state of the art in water system cybersecurity. The simulator and dataset produced represent a significant contribution to the research community, and we are now well-positioned to proceed with the development and validation of the next generation of intelligent, process-aware intrusion detection systems.

References

[1] N. Tariq, M. Asim, and F. A. Khan, "Securing scada-based critical infrastructures: Challenges and open issues," in *2023 International Conference on Communication, Computing and Digital Systems (C-CODE)*, pp. 1–8, 2023.

[2] M. Laiani, A. Tugnoli, and V. Cozzani, "Critical cybersecurity scenarios in drinking water treatment plants," in *2023 International Conference on Cyber-Physical Social Intelligence (ICCPSI)*, pp. 1–6, 2023.

[3] M. Alanazi, A. Mahmood, and M. J. M. Chowdhury, "Scada vulnerabilities and attacks: A review of the state-of-the-art and open issues," in *Proceedings of the 2023 Australasian Computer Science Week Multiconference*, pp. 1–10, 2023.

[4] N. d'Ambrosio, G. Capodagli, G. Perrone, and S. P. Romano, "Scass: Breaking into scada systems security," in *2023 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 306–311, 2023.

[5] A. Pinto, I. M. Pires, A. García-de Prado, F. J. Espinosa, and L. García, "Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure," in *World Conference on Information Systems and Technologies*, pp. 498–510, Springer, 2023.

[6] C. M. Davis, J. E. Tate, H. Okhravi, C. Grier, T. J. Overbye, and D. Nicol, "Scada cyber security testbed development," in *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 1313–1318, 2023.

[7] M. A. Teixeira, T. Salman, M. Zolanvari, R. Jain, N. Meskin, and M. Samaka, "Scada system testbed for cybersecurity research using machine learning approach," in *2023 IEEE Globecom Workshops (GC Wkshps)*, pp. 1095–1100, 2023.

[8] M. Alanazi, A. Mahmood, and M. J. M. Chowdhury, "Ics-ltu2022: A dataset for ics vulnerabilities," in *Proceedings of the 2023 Australasian Computer Science Week Multiconference*, pp. 1–5, 2023.

[9] A. Mughaid, S. Alzu'bi, A. A. A. Alkhatib, A. AlZioud, A. Al Ghazo, and I. AL-Aiasha, "Simulation-based framework for authenticating scada systems and cyber threat security in edge-based autonomous environments," in *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0252–0258, 2023.

[10] J. T. Meier, T. D. Nguyen, and N. C. Rowe, "Hardening honeypots for industrial control systems," in *2023 IEEE Conference on Communications and Network Security (CNS)*, pp. 372–380, 2023.

[11] A. V. Serbanescu, S. Obermeier, and D.-Y. Yu, "A flexible architecture for industrial control system honeypots," in *2023 IEEE 1st International Conference on Advanced Networking and Communication (ADNETCOM)*, pp. 1–7, 2023.

[12] V. L. Do, L. Fillatre, and I. Nikiforov, "Sequential monitoring of scada systems against cyber/physical attacks," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 7523–7528, 2023.

[13] F. Moazeni and J. Khazaei, "Sequential false data injection cyberattacks in water distribution systems targeting storage tanks; a bi-level optimization model," *Journal of Water Resources Planning and Management*, vol. 149, no. 5, p. 04023018, 2023.

[14] A. A. Albustami and A. F. Taha, "Breaking the flow and the bank: Stealthy cyberattacks on water network hydraulics," *Water Research*, vol. 231, p. 119615, 2023.

[15] D. Giannubilo, T. Giorgeschi, M. Carminati, S. Zanero, and S. Longari, "A deep learning approach for false data injection attacks detection in smart water infrastructure," in *Proceedings of the 18th International Conference on Availability, Reliability and Security*, pp. 1–10, 2023.

[16] M. M. Aslam, A. Tufail, K.-H. Kim, R. A. A. H. M. Apong, and M. T. Raza, "A comprehensive study on cyber attacks in communication networks in water purification and distribution plants: Challenges, vulnerabilities, and future prospects," *Water*, vol. 15, no. 13, p. 2437, 2023.

[17] M. N. K. Sikder, M. B. T. Nguyen, E. D. Elliott, and F. A. Batarseh, "Deep h2o: Cyber attacks detection in water distribution systems using deep learning," in *2023 IEEE International Conference on Big Data (Big Data)*, pp. 6156–6161, 2023.

[18] A. O. Khadidos, A. O. Khadidos, S. Selvarajan, *et al.*, "Cybersentry: Enhancing scada security through advanced deep learning and optimization strategies," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 1, pp. 101–112, 2023.

[19] P. Botta, P. Khare, and M. J. B. Reddy, "Real-time cyberattack detection for scada in power system," in *2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, pp. 1–6, 2023.

[20] G. Birihanu and I. Lendák, "Explainable correlation-based anomaly detection for ics," in *2025 IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 000–000, 2025.

[21] N. Raza and F. Moazeni, "Optimal cybersecurity framework for smart water system: Detection, localization and severity assessment," *Computers & Security*, vol. 126, p. 103061, 2023.

[22] D.-H. Kim, I.-Y. Noh, H.-Y. Choi, B.-H. Kang, and Y.-J. Lee, "A study on performance metrics for anomaly detection based on industrial control system operation data," *Electronics*, vol. 11, no. 1, p. 148, 2022.

[23] S. S. Patlolla, "An approach for formal analysis of the security of a water treatment testbed," Master's thesis, Iowa State University, 2017.

Appendix: Predicted Timeline Chart

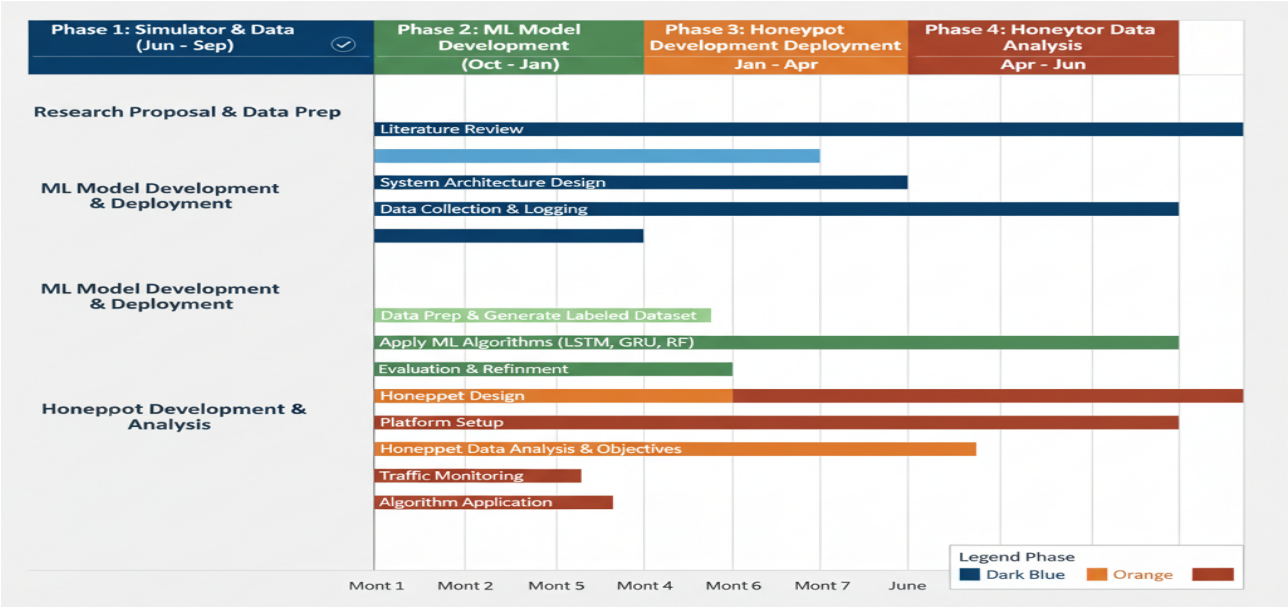

Figure 6.1: Predicted Timeline Chart