PlantShield: An Automated Response Tool for Critical Infrastructures

Sanskar Shrivastava, Tiger Lim, Aditya P. Mathur
iTrust, Centre for Cyber Security Research
Singapore University of Technology and Design Singapore
Email: sanskar_shrivastava@sutd.edu.sg, tiger_lim@mymail.sutd.edu.sg, aditya_mathur@sutd.edu.sg

Abstract-Cyber physical attacks on an Industrial Control System in a critical infrastructure could disrupt national economy and have far-reaching societal impacts globally. While anomaly detectors are useful in defending against cyberattacks, they fall short in the face of ransomware. During such attacks, engineers often experience partial system visibility, hindering their ability to manually control the plant. This impairment escalates risks to higher management levels. Additionally, the uncertainty of whether the adversary can control the plant or not is merely attacking the business logic that can prompt management to take drastic measures, such as shutting down the plant, which in turn affects dependent systems. This paper addresses these critical issues and proposes a solution. The proposed solution helps gain back system control when under an attack. Even when the attacker has gained access to plant controls, they cannot damage plant components. A version of the solution has been implemented in an operational water treatment plant and validated against typical cyber physical attacks.

Keywords: Cyber physical attacks, critical infrastructures, Industrial Control System, water treatment plant

1. Introduction

Cyber-Physical Systems (CPS) encompass systems operating within both the cyber and physical domains. Examples include Intelligent Transportation Systems, Water Treatment and Distribution Systems, automobiles, pacemakers, and the Power Grid. CPS integrates computing, networking, and physical processes to create tangible impacts in the physical world[1]. Computers monitor, control, and integrate these physical processes, with the intelligence in the cyber domain determining the actions based on the system's state. This integration facilitates automation, control, and quality assurance, reducing the need for human intervention. Consequently, cybersecurity becomes crucial for CPS, as Critical Infrastructure (CI) rely on this model. The security of CPS involves more complexities than traditional networked Cyber Systems. A secure CPS has at least the following requirements

- Confidentiality,
- Integrity,
- Availability, and
- Authenticity.

The risk extends beyond data and services to include physical and biological threats to individuals impacted by the CPS. Since adversaries can physically attack the system's processes, defense mechanisms must also account for physical security.

2. Cyber Physical Attacks and Critical Infrastructures

2.1. What are Cyber Physical Attacks

Due to distribution of physical processes and communications, increased network connectivity, and technology readiness, contemporary National Infrastructures face the following threats[2]:

- Cyberattacks – an attack exclusively in the cyber domain, where confidentiality, integrity, and availability are compromised.

979-8-3315-2096-0/25/$31.00 ©2025 IEEE

- Physical attacks – attacks purely in the physical domain.
- Cyber physical attacks – attacks in the cyber domain with consequences in the physical domain such as unauthorized actuation, espionage, and delayed actuation.
- Physical cyberattacks – attacks on the physical domain with consequences in the cyber domain.

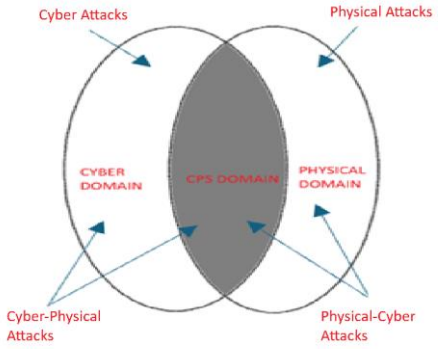

Fig 2.1 Cyber-Physical Convergence

2.2. How Cyber Physical Attacks affect Critical Infrastructures

The CIA triad is a fundamental concept in information security, representing three core principles: Confidentiality, Integrity, and Availability. Compromising any of these element results in cyber impacts as well as physical impacts. This may include:

- Gaining credentials allows an attacker to circumvent authorization mechanisms.
- Interfering with communication can render the system unavailable.
- Problems like delayed and unauthorized actuation.

Cyber-physical attacks[6] on critical infra-structure can have severe and far-reaching impact. These attacks target the digital systems that control physical processes, leading to disruptions in essential services and potential physical damage. Here are some key effects:

1. Service Disruptions: Attacks can lead to blackouts, water supply interruptions, and transportation system failures. For instance, the 2021 Colonial Pipeline attack caused fuel shortages throughout the Eastern United States.
2. Economic Impact: These disruptions can result in significant economic losses.

Other significant effects include National Security Threats, physical damage to CI. Therefore, there is an increasing emphasis on enhancing the security of these systems to mitigate these risks.

2.3 Secure Water Treatment Testbed

The Secure Water Treatment (SWaT) testbed , used in our research, consists of a six-stage process. The process begins by taking in raw water, adding necessary chemicals, filtering it using an ultrafiltration (UF) system, de-chlorinating it via ultraviolet lamps, and feeding it to a reverse osmosis (RO) system. A backwash process cleans the membranes in the UF system using the rejected water from the RO system. The cyber portion of the SWaT testbed consists of a layered communications network, programmable logic controllers (PLCs), human-machine interfaces (HMIs) (as shown in Figure 6), supervisory control and data acquisition (SCADA) workstation, and a historian system [3][5]

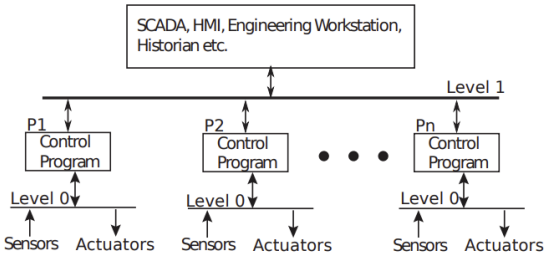

Fig 2.3.1 Architecture of the control portion of a CPS

3. Proposed Method

3.1. PlantShield

PlantShield is a software-hardware subsystem designed to monitor and control the security, safety, and service of an Industrial Control System (ICS) or critical cyber-physical infrastructure. PlantShield is built as a prototype for new “secured by design” critical infrastructure. A version of PlantShield has been demonstrated in the SWaT. In conjunction with other tools and cyber twins, PlantShield can be used for training, teleoperation, automatic incidence response, and visual forensics to assist the cyber responder in all stages of the National Institute of Standards and Technology

(NIST) Cybersecurity Framework which are Identify, Protect, Detect, Respond and Recover.

3.2 Working of PlantShield

The current version of PlantShield has been built using a Raspberry Pi 4 Model B computer. It connects to PLC and RIO through ethernet ports on the Raspberry Pi and PLC. Hence, any communication between PLC and RIO can be intercepted with the help of the Pi. Plant Shield has works in three key steps listed next.

1) State Validation: In this step, the code hosted in PI validates the current state of the actuator, e.g., a pump, against what it should be. If the states observed and expected match, validation is considered successful and not otherwise.
2) Command Correction: When validation fails, the Pi generates the correct command to be issued to the actuator.
3) Command Transmission: The corrected command is then transmitted. The PI captures the incoming Digital Output (DO) packet from the PLC, replaces the incorrect DO command with the correct one, and transmits it to the RIO.

Thus, the actuator remains in good state despite the attack that resulted in a “bad command” sent by the PLC to the actuator.

4. Experiments and Result

The following experiments demonstrate the effectiveness of PlantShield in different scenarios for stage 1 of SWaT.

Experiment 1: Under Normal Conditions, i.e., the plant is running with no attack.

Result: As we can see from Fig 4.1, no invalid commands were reported by PlantShield. It was tested at random intervals.

Experiment 2: Spoofing Attack on Tank Level

A spoofing attack was launched on level sensor LIT101 for tank T101. This caused LIT101 to b set to a LOW level on SCADA despite actually being HIGH. This resulted in opening of valve MV101 which is responsible for filling water in T101. If the attack were allowed to continue, water in T101 would overflow. PlantShield recognizes from level zero network data that MV101 should be closed and sends the correct command to do so.

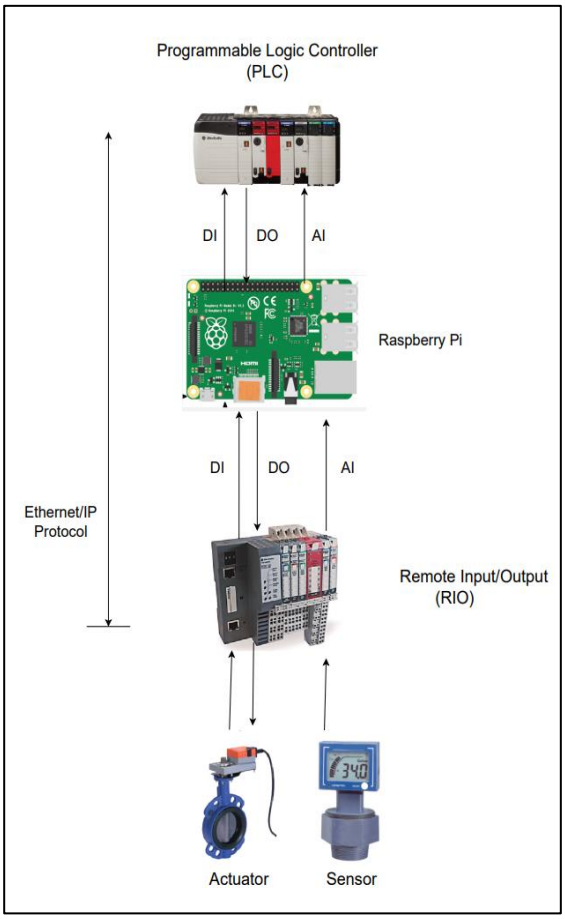

Fig. 3.2.1 Communication of PlantShield (programmed in RPi) with PLC and RIO

.

```
PLANT PROTECT LOADED

Collecting L0 Data

0) LIT101 is 915.806043322  LIT301 is 851.063829787   P101 is 0   MV101 is 0   MV201 is 0
1) LIT101 is 915.806043322  LIT301 is 851.063829787   P101 is 0   MV101 is 0   MV201 is 0
2) LIT101 is 915.806043322  LIT301 is 851.063829787   P101 is 0   MV101 is 0   MV201 is 0
3) LIT101 is 916.120065368  LIT301 is 851.063829787   P101 is 0   MV101 is 0   MV201 is 0
4) LIT101 is 916.120065368  LIT301 is 851.063829787   P101 is 0   MV101 is 0   MV201 is 0
5) LIT101 is 916.120065368  LIT301 is 851.664637272   P101 is 0   MV101 is 0   MV201 is 0
6) LIT101 is 915.570526788  LIT301 is 851.945014099   P101 is 0   MV101 is 0   MV201 is 0
7) LIT101 is 915.570526788  LIT301 is 851.945014099   P101 is 0   MV101 is 0   MV201 is 0
8) LIT101 is 915.020988208  LIT301 is 851.945014099   P101 is 0   MV101 is 0   MV201 is 0
9) LIT101 is 915.020988208  LIT301 is 852.145283261   P101 is 0   MV101 is 0   MV201 is 0
10) LIT101 is 915.177999231  LIT301 is 852.145283261   P101 is 0   MV101 is 0   MV201 is 0
11) LIT101 is 915.177999231  LIT301 is 852.145283261   P101 is 0   MV101 is 0   MV201 is 0
12) LIT101 is 914.628460651  LIT301 is 852.145283261   P101 is 0   MV101 is 0   MV201 is 0
13) LIT101 is 914.628460651  LIT301 is 852.145283261   P101 is 0   MV101 is 0   MV201 is 0
14) LIT101 is 914.628460651  LIT301 is 852.065175596   P101 is 0   MV101 is 0   MV201 is 0
15) LIT101 is 914.628460651  LIT301 is 851.344206614   P101 is 0   MV101 is 0   MV201 is 0
16) LIT101 is 914.628460651  LIT301 is 851.344206614   P101 is 0   MV101 is 0   MV201 is 0
17) LIT101 is 914.432196873  LIT301 is 851.344206614   P101 is 0   MV101 is 0   MV201 is 0
18) LIT101 is 914.432196873  LIT301 is 851.344206614   P101 is 0   MV101 is 0   MV201 is 0
19) LIT101 is 914.432196873  LIT301 is 851.344206614   P101 is 0   MV101 is 0   MV201 is 0

Starting Validation

NO change in LIT101
NO change in LIT301

Current state of P101 is 0
Correct state should be 0
Current state of MV101 is 0
correct state of MV101 is 0

Command Valid for P101

Command Valid for MV101

No suspicious activity observed.

Data Collection to start in 5 seconds!
```

Fig 4.1 Experiment 1 Result

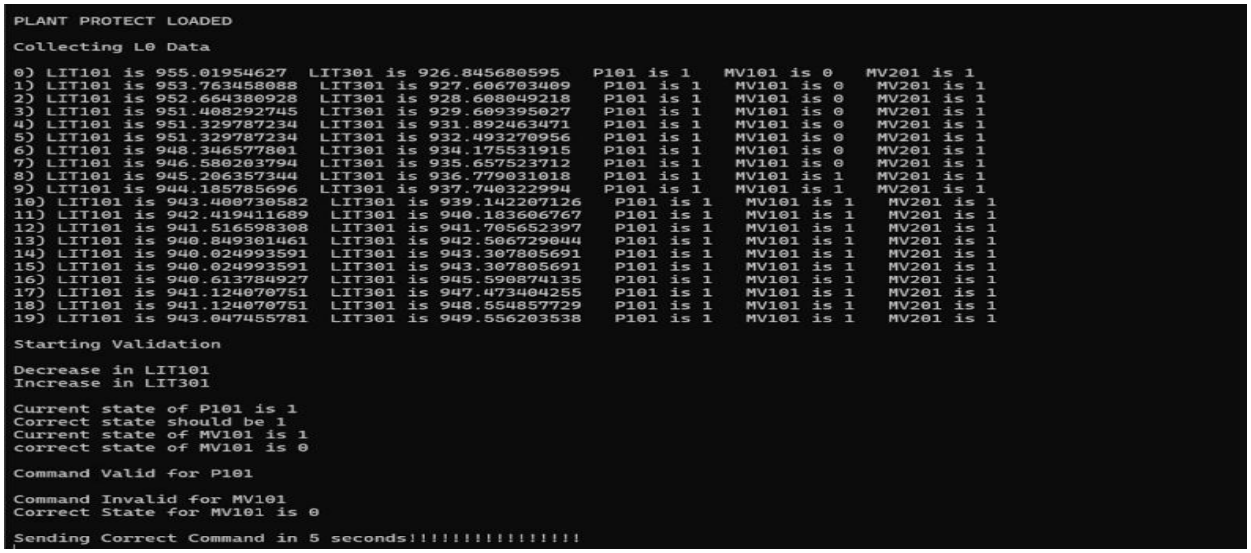

Fig 4.2 Experiment 2 Result

Experiment 2 Result: As can be seen from output of the PlantShield that MV101 state is 1 (OPEN) and the correct state should be 0 (CLOSED). Hence it sends a CLOSE command during command transmission. (Fig 4.2)

Experiment 3: Attack on an actuator
In this attack P101 was forced to run while valve MV201 was open. If water level in LIT301 is high, it would eventually lead to flooding if P101 is not stopped in time. With PlantShield active, it detected P101 is running when it should not be. Here P101 is in invalid state and the correct command was transmitted by PlantShield before any flooding.

Experiment 3 Result: Plant Shield sends the command to stop P101.(fig 4.3)

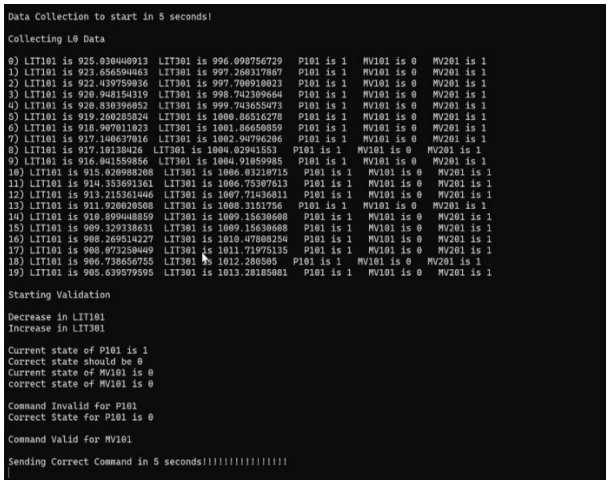

Fig 4.3 Experiment 3 Result

5. Conclusion

PlantShield demonstrated its effectiveness by successfully taking control of the actuators under attack, ensuring the correct commands were sent to maintain system integrity. During normal operations, PlantShield did not detect any anomalies or issue incorrect commands, indicating its reliability in stable conditions. Even amidst an attack, PlantShield occasionally reported valid commands, as the sensor values and plant state remained within acceptable predefined levels, preventing any damage. This resilience highlights the PlantShield's ability to discern between harmful and non-harmful situations. In instances where the plant's state deviated from expected values due to an attack, PlantShield accurately identified the invalid state. It then intervened by sending the correct command to the actuator, thereby mitigating potential risks and ensuring the system's continued safe operation.

6. References

[1] Shrivastava Siddhant, Sridhar Adepu, and Aditya Mathur. "Design and assessment of an orthogonal defense mechanism for a water treatment facility." Robotics and Autonomous Systems 101 (2018): 114-125.

[2] Francisco Furtado; Mark Goh; Siddhant Shrivastava; Aditya Mathur, "The Design of Cyber-Physical Exercises (CPXS)." 2022 14th

International Conference on Cyber Conflict (CyCon). Vol. 700. IEEE, 2022

[3] Urbina, David and Giraldo, Jairo and Tippenhauer, Nils Ole and Cardenas, Alvaro, “Attacking fieldbus communications in ICS: Applications to the SWaT testbed,” IOS Press, 2016.

[4] Giedre Sabaliauskaite, Aditya P. Mathur,” Intelligent Checkers to Improve Attack Detection in Cyber Physical Systems”, 2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery

[5] Aditya P. Mathur, Nils Ole Tippenhauer, ”SWaT: a water treatment testbed for research and training on ICS security” 2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)

[6] Sridhar Adepu, Aditya P. Mathur, ”Distributed Attack Detection in a Water Treatment Plant: Method and Case Study”2018, IEEE Transactions on Dependable and Secure Computing”

[7] Jonathan Goh; Sridhar Adepu; Marcus Tan; Zi Shan Lee,” Anomaly Detection in Cyber Physical Systems Using Recurrent Neural Networks”, 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE)