

UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Computer Science
Software Engineering Curriculum

Ojus Virendra Tudavekar

Blockchain and Digital Twin-based Approach for Securing Water Supply Infrastructure

Master's Thesis (30 ECTS)

Supervisor: Mubashar Iqbal, PhD

Tartu 2024

Blockchain and Digital Twin-based Approach for Securing Water Supply Infrastructure

Abstract:

Digital Twin (DT), as a virtual representation of physical entities, provides valuable insights into the Cyber-Physical System's (CPS's) behavior and characteristics. DT's capabilities of monitoring, visualizing, testing, and simulating the physical process have been widely used in industrial systems over the years to improve quality and efficiency. Moreover, in this era of increasing digitization, the convergence of water infrastructure and cybersecurity emerges as a critical concern. DT, which is usually seen as a benchmark for generating virtual replicas of real-world objects, holds significant potential in enhancing the security and resilience of the Water CPS. Integrating blockchain and DT technology for industrial systems has recently gained popularity among researchers. There is a dearth of research on using DT to enhance security in Water CPS. In this work, we present an extensive literature review of existing literature on Water CPS where primary security threats, vulnerabilities exploited, and proposed detection mechanisms are analyzed. Moreover, a novel approach of DT integrated with blockchain as an attack deception mechanism is proposed to enhance the security of Water CPS, using the Secure Water Treatment (SWaT) system as a base CPS architecture. Additionally, the attacker models, attack scenarios, and role-based Incident Response Playbooks (IRPs) to contain and mitigate the attacks in Water CPS are showcased. The proposed solution is evaluated using the role-based IRP for various attack scenarios and DT-based simulation with Microsoft Azure Digital Twin Platform.

Keywords:

Digital twin, Blockchain, Cyber-Physical System, Cybersecurity, Deception, Incident Response Playbook.

CERCS: P170, Computer Science, Numerical Analysis, Systems, Control

Blockchaini ja digitaalse kaksiku põhinev lähenemisviis veevarustuse infrastruktuuri kindlustamiseks

Lühikokkuvõte:

Digitaalne kaksik (inglise keeles Digital Twins ehk DTs) kui füüsiliste üksuste virtuaalne kujutis annab väärtusliku ülevaate küberfüüsilise süsteemi (CPS) käitumisest ja omadustest. DT võimeid jälgida, visualiseerida, testida ja simuleerida füüsilist protsessi on tööstussüsteemides aastate jooksul laialdaselt kasutatud kvaliteedi ja tõhususe parandamiseks. Lisaks sellele on praegusel kasvava digitaliseerimise ajastul kriitilise tähtsusega vee infrastruktuuri ja küberturvalisuse lähenemine. DT, mida tavaliselt peetakse reaalsete objektide virtuaalsete koopiade loomiseks, omab märkimisväärset potentsiaali veemajanduse CPSi turvalisuse ja vastupidavuse suurendamisel. Plokihela ja DT-tehnoloogia

integreerimine tööstussüsteemide jaoks on hiljuti teadlaste seas populaarsust kogunud. Teadusuuringuid DT kasutamise kohta vee- ja ühissüsteemi turvalisuse suurendamiseks on vähe. Käesolevas töös esitame ulatusliku kirjanduse ülevaate olemasolevast kirjandusest vee CPSi kohta, kus analüüsitakse esmaseid julgeolekuohte, kasutatavaid haavatavusi ja pakutud avastamismehhanisme. Lisaks sellele pakutakse välja uudne lähenemisviis, mille puhul DT on integreeritud plokiahelaga kui rünnaku pettusmehhanismiga, et suurendada veekäitluskeskuste turvalisust, kasutades turvalise veepuhastussüsteemi (SWaT) kui veekäitluskeskuste baasarhitektuuri. Lisaks tutvustatakse ründajate mudeleid, ründestsenaariume ja rollipõhist intsidentidele reageerimise käsiraamatut (IRPs), et ohjeldada ja leevendada rünnakuid Water CPSis. Kavandatud lahendust hinnatakse, kasutades rollipõhist IRP-d erinevate ründestsenaariumide jaoks ja DT-põhist simulatsiooni Microsoft Azure Digital Twin Platform abil.

Võtmesõnad:

Digitaalne kaksik, plokiahel, küberfüüsiline süsteem, küberturvalisus, pettus, intsidentidele reageerimise mängukiri.

CERCS: P170, Arvutiteadus, arvanalüüs, süsteemid, kontroll

Acknowledgement

I'd like to first thank my supervisor, Dr. Mubashar Iqbal for patiently answering all my questions related to this work and being the kind guiding force needed for the project.

I'd also like to thank Ashfaq Hussain Ahmed for helping me learn the core concepts of implementing this work and being the best support.

I thank my parents, Virendra and Deepali, my sister Rajalaxmi, and my brother Vikrant for being my pillars of support.

I thank my friends PV Varun, Rasinthe, and Ameer Hamza for keeping me company through thick and thin.

Lastly, I'd like to thank my friend Kairit Kruusa for the Estonian translations of the abstract and for being the biggest motivation through this journey.

Cyber-security Excellence Hub in Estonia and South Moravia

This work is part of the Cyber-security Excellence Hub in Estonia and South Moravia (CHESS: <https://chess-eu.cs.ut.ee>) project funded by the European Union under Grant Agreement No. 101087529. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Grammarly

Throughout the composition of this work, the digital writing assistant, Grammarly¹, served as an invaluable tool. Grammarly, an AI-powered software, is specifically designed to aid users in refining their writing skills by providing suggestions for grammar, spelling, punctuation, clarity, and style. Accessible via a web-based service, browser extension, and desktop application compatible with Windows and Mac operating systems, Grammarly offers comprehensive support for various types of writing, including emails, essays, and reports. Through sophisticated algorithms and natural language processing capabilities, Grammarly delivers real-time suggestions, enabling users to promptly address errors and enhance the overall quality of their writing. Key features of Grammarly encompass spell and grammar checks, punctuation recommendations, readability and clarity analysis, and vocabulary enhancement.

ChatGPT

¹<https://www.grammarly.com/>

In the process of composing this work, ChatGPT [1] version V3.5 has been utilized. Developed by OpenAI², ChatGPT is an advanced AI language model founded on the GPT (Generative Pre-trained Transformer) architecture, notably GPT-3.5. It functions as a brainstorming assistant, proficient in comprehending and generating human-like text based on input stimuli. Leveraging extensive pre-training on vast datasets sourced from the internet, ChatGPT has acquired proficiency in grammar, vocabulary, and diverse linguistic structures. Its adaptable nature allows for fine-tuning to specific tasks, rendering it versatile across a broad spectrum of applications. ChatGPT finds utility in diverse domains, including customer support, content creation, programming assistance, language translation, and engaging in interactive dialogues with users.

²<https://chat.openai.com/>

Contents

1	Introduction	11
1.1	Motivation	12
1.2	Problem statement	12
1.3	Research Questions	13
1.4	Contributions	13
1.5	Thesis Structure	14
2	Background	15
2.1	Cyber Physical System	15
2.2	Secure Water Treatment (SWaT) Testbed	16
2.3	Deception	19
2.4	Digital Twin	21
2.4.1	Architecture and Components	22
2.4.2	Digital Twin Advantages Over Traditional Honeypot	23
2.5	Security Operations Center (SOC) Playbook	25
2.6	Blockchain	26
2.6.1	Leveraging the Data Immutability feature of Blockchain	27
2.6.2	Types Of Blockchain	28
2.6.3	Ethereum	28
2.6.4	Hyperledger Fabric	29
2.6.5	Why Ethereum?	30
3	Systematic Literature Review	30
3.1	Review Questions	31
3.2	Review Settings	31
3.2.1	Search Strings	32
3.2.2	Data Sources	32
3.2.3	Inclusion and Exclusion Criteria	33
3.2.4	Paper Selection	33
3.2.5	Data Extraction Strategy	34
3.3	Presentation of Results	34
3.3.1	Water systems: Challenges and Operations	35
3.4	Identified gaps	48
3.4.1	Challenges in the Proposed Current System	49
3.4.2	Why Digital Twin Technology?	49
3.4.3	Why Blockchain?	50
3.5	Summary	51

4	Use Case and Attack Scenarios	52
4.1	Use Case 1: Water Supply Infrastructure and Cybersecurity	53
4.2	Use Case 2: Digital Twins in Water Supply	54
4.3	Attacker Model and Scenarios	55
4.3.1	Attacks Targeting P1 stage	57
4.3.2	Attack scenarios	57
4.4	Summary	59
5	Solution Design	59
5.1	Proposed HLC Portion Architecture Of SWaT With Digital Twin	60
5.2	Proposed Digital Twin Design	60
5.3	Ethereum Based Solidity Smart Contract	62
5.4	SWaT Dataset	63
5.5	Anomaly Detection System	64
5.6	Proposed Incident Response Playbook Design	64
5.7	Summary	65
6	Implementation	66
6.1	Proposed Attack Deception Architecture Of SWaT with Digital Twin . .	66
6.2	Digital Twin Models	68
6.2.1	Flow Sensor: FIT101	69
6.2.2	Flow Sensor: FIT201	69
6.2.3	Level Sensor: LIT101	70
6.2.4	Motorized Valve: MV101	70
6.2.5	Pump: P101	71
6.2.6	Tank: T101	72
6.3	Digital Twin Simulators	73
6.4	Digital Twin Data Ingestors	78
6.5	Solidity Smart Contracts	83
6.5.1	Tank T101 Contract	85
6.5.2	Level Sensor LIT101 Contract	85
6.5.3	Process Stage Contract	86
6.5.4	Deployment Script	87
6.6	Azure Digital Twin Explorer	88
6.7	Summary	89
7	Evaluation	90
7.1	Incident Response Playbook Implementation	90
7.2	Summary	95

8	Discussion	96
8.1	Answers to Research Questions	96
8.2	Limitations	97
8.3	Future Work	98
9	Conclusion	99
	References	103
	Appendix	104
	I. Resources	104
	II. Licence	105

List of Figures

1	Typical Cyber-Physical System adapted from [40]	15
2	SWaT Process Diagram. P1-P6 indicate the six stages in the treatment process. Solid arrows indicate flow of water or chemicals in the dosing station. Dashed arrows indicate potential cyber attack points. LIT : Level Indicator and Transmitter; Pxxx : Pump; AITxxx : Property indicator and Transmitter; FITxxx : flow meter; DPIT : Differential Pressure Indicator and Transmitter.	17
3	Control Portion Architecture of a CPS.P1, P2,. . . ,Pn denote PLCs. SW1 and SW2 are Switches. Each PLC communicates with its sensors and actuators through a local network at Level 0. PLCs communicate among themselves via another network at Level 1. Communication with SCADA and other computers is via a Level 3 network not shown here. Note that the actuators, e.g., a pump, also have sensors to indicate their condition, adapted from [8].	18
4	Example architecture for before and after deployment of deception, adapted from [27].	20
5	3D DT Model Representation, adapted from [16]	22
6	5D DT Model Representation, adapted from [32]	23
7	Paper Selection Process	34
8	Proposed High-Level Control Portion Architecture of SwaT with DT as Deception Technique.C1 denotes communication with DT in Level 2. ADS denotes Anomaly Detection system and DE denotes broken connection for fake Data exchange between physical Twin and Digital Twin	61
9	P1 Stage Process of SWaT	62
10	Implementation of Proposed Attack Deception Architecture	67
11	Instances Of Device Components in Azure IOT Hub	73
12	Level sensor simulator sending telemetry data	77
13	Event subscriptions in Azure IOT Hub for the devices	78
14	Azure function for the Level Sensor LIT101 in Azure IOT Hub	79
15	Data Ingestion by Azure Function Flow Diagram	80
16	Verification of LIT101 Property update on ADT Explorer	84
17	Verifying LevelSensor LIT101 DT update on ADTEplorer	84
18	Successful Deployment of contracts showing in Etherscan	88
19	Digital Twins with relationships in ADT Explorer	89
20	Incident Response Playbook With DT As Attack Deception. Numbers 1-15 denote the Steps.	91

List of Tables

1	Advantages of Digital Twins over traditional Honeypots as Attack De-ception Technique	24
2	Data Extraction Form	35
3	Past Cyber-attacks on Water Systems	36
4	Past Incidents Analysis	40
5	Detection methods	44
6	Attacks targeting P1 Stage process of SWaT	57
7	Mapping of Research Questions with Work	97

1 Introduction

Water systems serve as indispensable infrastructures crucial for delivering safe and clean water to communities [40]. The water systems face many challenges in their efforts to continue providing services, including water pollution, rising urbanization and population growth, ineffective infrastructure, and adherence to stricter regulations and water quality standards [40]. Water and wastewater providers are embracing smart water systems that are dependable and effective and facilitate real-time decision-making in order to address these issues [40]. Water systems are a kind of Cyber-Physical systems (CPS) that integrate computational and physical capabilities to control and monitor physical processes [40]. Key components of existing water systems, such as Programmable Logic Controllers (PLCs) and Supervisory Control and Data Acquisition (SCADA) systems, have already embraced CPS. PLCs facilitate communication for actuator control, bridging sensors, and actuators, while SCADA systems supervise entire water infrastructures, storing and analyzing real-time data [23]. However, the interconnectivity inherent in these systems exposes them to cyber threats, jeopardizing water supply, quality, and public health. The interconnected nature of these systems also makes them vulnerable to cyber attacks that could disrupt water supply, compromise quality, and pose health risks [40]. Notable incidents such as Stuxnet [12], DuQu [33], BlackEnergy [20], and Havex [11] serve as stark reminders of the catastrophic consequences cyberattacks can inflict.

Water systems security mainly relied on isolation and restricted access to control components, utilizing primarily physical devices like pumps, valves, and pipes [40]. However, the emergence of the Internet of Things (IoT) and smart system principles is revolutionizing water systems and other critical infrastructures [40]. This shift towards Industry 4.0 entails integrating IoT and analytics into industrial control systems to enhance sensing and control capabilities and ensure better integration with business processes. While offering optimization and predictive maintenance opportunities, this technological evolution also presents various security challenges [40]. Various security threats [5](also discussed in Section 3) to water systems revealed a lack of security mechanisms and defense strategies as the primary security challenges in water systems.

In this context, this work aims to define a novel approach of Digital Twin (DT) integrated with Blockchain as an attack deception mechanism to enhance the security of Water CPS using the Secure Water Treatment (SWaT) system as a base CPS architecture. We also delve into incident response strategies and methodologies to investigate and mitigate security threats to water systems. With a specific focus on water systems, this research contributes to preserving a vital resource for communities and lays the groundwork for broader insights into DT and Blockchain applications in securing critical infrastructure.

1.1 Motivation

The imperative to secure water systems, fundamental to sustaining life and preserving public health, underpins the core motivation of this study. As digital technologies increasingly interlace with physical infrastructure, the development of innovative solutions to shield water systems from cyber threats emerges as a pressing necessity. The integration of network communication, the proliferation of Commercial-Off-The-Shelf (COTS) components, and the widespread deployment of wireless systems within Purdue and SWAN architecture layers introduce new security challenges, potentially exposing water systems to diverse adversaries [5]. The rising frequency of reported attacks targeting CPSs crucial for national infrastructure services underscores the urgency of addressing these vulnerabilities. Cyber attacks on infrastructure services are frequently not made public, and identifying the people responsible for these occurrences can be difficult and unpredictable, requiring a high level of expertise [40]. Still, publicly disclosed incidents such as Maroochy Water Services 2000 [31] and Pennsylvania Water Filtering Plant 2006 [36] indicate that a broad range of entities appears to be the origins of cyberattacks targeting water infrastructure. Water systems provide an appealing target for political, military, and terrorist actors alike, given the importance of water to ecological balance and human existence [40]. Recent studies have highlighted the vulnerabilities of water systems to cyberattacks, with notable incidents such as Stuxnet, DuQu, BlackEnergy, and Havex serving as stark reminders of the catastrophic consequences such attacks can inflict[40].

As cyber threats evolve and cyberattacks on critical infrastructures become increasingly prevalent, there is a compelling need to explore innovative approaches to fortify water systems against emerging risks. The convergence of water infrastructure and cybersecurity presents a unique opportunity to leverage technologies like DT and Blockchain for enhanced resilience. However, the existing literature lacks comprehensive research dedicated explicitly to the utilization of DTs in bolstering the cybersecurity of water systems. By bridging this research gap and investigating the potential applications of DT technology in enhancing water system security, this work aims to contribute valuable insights and methodologies to secure critical infrastructure in an increasingly digitized world. The proposed research will not only enhance our understanding of the vulnerabilities inherent in water systems but also provide practical solutions and strategies to mitigate cyber threats effectively. This interdisciplinary approach aligns with the growing emphasis on cybersecurity in critical infrastructure sectors and underscores the importance of proactive measures to secure essential services for society.

1.2 Problem statement

CPS, particularly in the domain of water infrastructure, faces significant cybersecurity challenges that threaten its integrity, reliability, and safety. These challenges encompass

a wide range of threats, from unauthorized access to data tampering, which can have profound implications for public health and environmental sustainability as observed in incidents such as Maroochy Water Services 2000 [31] and Pennsylvania Water Filtering Plant 2006 [36]. Despite implementing conventional security measures, CPS in the water sector remains vulnerable to increasingly sophisticated cyberattacks. Lack of innovation and technology in the domain of water system security was revealed to be the main concern. The integration of emerging technologies such as DT and blockchain holds great potential for bolstering the security and resilience of water CPS. However, a notable gap exists in understanding how these innovative technologies can be effectively harnessed to mitigate cybersecurity risks within the context of water infrastructure. Furthermore, the absence of comprehensive incident response frameworks tailored specifically to the unique characteristics of water CPS exacerbates the challenges of securing these critical systems.

Therefore, the primary aim of this work is to address these pressing challenges by proposing a novel approach that leverages DT and blockchain technology to enhance the security posture of Water CPS. Additionally, this research endeavors to develop and validate a robust incident response playbook customized to the distinct requirements of water CPS. This work seeks to enable effective detection, response, and mitigation of cybersecurity incidents within water infrastructure by empowering operators and security personnel with tailored tools and strategies.

1.3 Research Questions

To address the aforesaid research problems, we formulate the main research question *"How can Digital Twins be used to Enhance the Cybersecurity of Water CPS?"* To answer this research question, we prepared four sub-research questions:

- **RQ1:** What role can DT play in improving the security posture of Water CPS?
- **RQ2:** What architectural frameworks facilitate the integration of DTs into Water CPS security?
- **RQ3:** How can Blockchain technology enhance the security of DTs in Water CPS?
- **RQ4:** How can an Incident Response Playbook be created to enhance the incident response strategy in Water CPS?

1.4 Contributions

This work contributes to the field of water system cybersecurity by leveraging DT, Blockchain, and incident response strategies. Firstly, it investigates the pivotal role of DTs in fortifying the security posture of Water CPS, elucidating their potential to enhance

system resilience and mitigate cyber threats. Secondly, it explores architectural frameworks tailored for the seamless integration of DTs into water CPS security infrastructure, aiming to optimize system performance while ensuring robust cybersecurity measures. Moreover, the work delves into the transformative potential of Blockchain technology in securing the integrity and authenticity of DT data. Through a systematic literature review, it scrutinizes existing research on securing industrial systems, with a specific focus on DT and Blockchain applications. This comprehensive review provides valuable insights into emerging trends and best practices in leveraging DT and Blockchain to address cybersecurity challenges in critical infrastructure. Lastly, the work develops tailored Incident Response Playbooks (IRPs) designed specifically for Water CPS environments. These playbooks equip stakeholders with effective strategies to detect, respond to, and recover from cyber incidents swiftly and efficiently. By combining these innovative approaches, this research aims to advance the cybersecurity resilience of water systems, ensuring the continuous delivery of safe and reliable water services to communities.

1.5 Thesis Structure

The further work is organized as follows. Section 2 presents the details of the main concepts a reader might encounter in this work. Details of CPS and DT with their definitions, architecture, components, and applications. Deception, the role of blockchain in the data security of DT, the significance of the security operations center playbook in incident response, and a detailed overview of the Security Water Treatment testbed (SWaT) are also explained. Section 3 maps out a systematic literature review process on existing literature where primary security threats to water systems, Vulnerabilities exploited, and detection mechanisms proposed were analyzed. This section also discusses the identified gaps and challenges in the existing literature and how DT and blockchain technology can address the gaps. Section 4 exhibits the Use case regarding the water supply infrastructure and cybersecurity and the role of DT's in the water supply infrastructure. This section also sheds light on the Attacker model and various attack scenarios built around the SWaT attack dataset to showcase the proposed solution's potential. Section 5 illustrates the design artifacts for implementing the proposed DT integrated with blockchain as an attack deception mechanism and IRP. Section 6 accumulates the implementation of the proposed solution and IRP. Section 7 showcases the implementation of IRP and the potential of the proposed solution by evaluation using the role-based IRP. Section 8 discusses the answers to the research questions of the work and the mapping of research questions with the work. This section also showcases the proposed solution's limitations and points out the future direction of this work. Finally, Section 9 concludes this work, mapping out this work's findings and summarizes the potential of the proposed DT integrated with Blockchain and IRP as an attack deception mechanism to enhance the security of Water CPS.

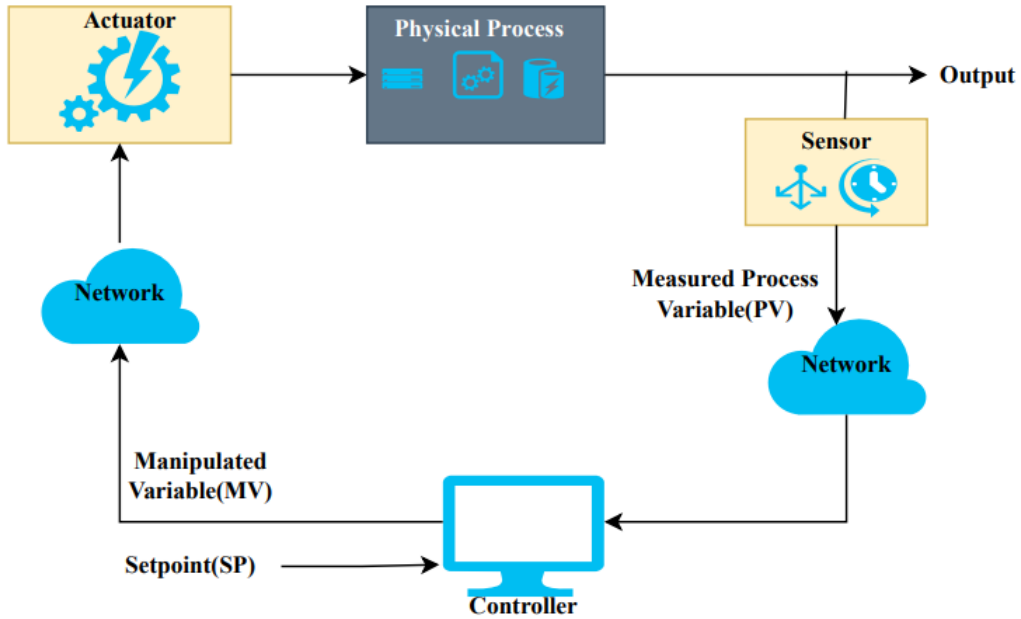


Figure 1. Typical Cyber-Physical System adapted from [40]

2 Background

This section defines the fundamentals of CPS and DT, clarifying their definitions, architectures, and components. It delves into the concept of deception as a defense mechanism, the role of blockchain in ensuring data security, the significance of Security Operations Center (SOC) playbooks for incident response, and offers a detailed overview of the SWaT testbed.

2.1 Cyber Physical System

CPS is found in a broad range of sectors, including healthcare, materials, manufacturing, automotive, aerospace, utilities, chemical, civil infrastructure, and transportation. Despite the differences in interpretation, many industry sectors share common technologies and, by extension, share similar concerns relating to their security [40]. A common concern for all these sectors in adopting new enabling technologies for CPS is to ensure security in the face of cyber-attacks. The figure 1 shows typical components of a networked CPS. We Can modify it and make it more detailed accordingly.

- The controller is given a process reference (Setpoint-SP) as the desired process output to maintain.

- The sensor measures the output of the physical process (Measured Process Value-PV) and sends this over a network to the controller.
- The controller (for example, a PLC) receives these values, compares the PV against the desired SP reference value, calculates a control command (Manipulated Variable-MV), and sends this through the network to the actuator.
- The actuator acts on this command and outputs a physical control action that modifies the process.

Although CPS's technology is efficient and based on information communication technology, it is vulnerable to cyber-attacks due to information disturbance of sensor and communication server [23]. As the model proposed in [40], attacks against CPS involve attacking components of CPS to achieve either data exfiltration, which involves gathering sensitive information about the CPS, or sabotage, which involves disrupting the process. Attackers use different types of tools to carry out the attack against elements of the CPS. The success of an attack depends on the resources and skills available to adversaries as well as system vulnerabilities and the absence of appropriate independent layers of protection designed to prevent mal-operation due to operator error, random equipment failure, or cyber-attack [40].

Smart Water Distribution System CPS: A Smart Water Distribution system can be defined as a water supply system upgraded with technologies such as sensing (via sensors and monitors), real-time communications (such as wireless networks, satellite communications, etc.), controls, and intelligence. The overview of the components of the Smart Water Distribution system is as follows [5]. The major components that form a Smart Water Distribution system include water tanks, pipes, smart water meters, smart pressure meters, flowmeters, energy consumption (pumping) meters, smart water treatment monitors, smart water purity sensors, physical security monitors, smart river height sensors, dam height sensors, levee movement sensors smart valves, smart pumps, smart contaminant sensors, smart flood sensors, etc [5]. All the mentioned smart components above can potentially be attacked along the systems used to connect them, Thus considering security factors of Smart Water Distribution systems become crucial.

2.2 Secure Water Treatment (SWaT) Testbed

The Secure Water Treatment (SWaT) [25] testbed serves as an experimental platform for water treatment, emulating the structure of a contemporary water treatment plant commonly found in urban settings. In a small footprint producing 5 gallons/minute of doubly filtered water, this testbed mimics large modern plants for water treatment, such as those found in cities [7]. It is designed and constructed to facilitate experimental investigations into the development of secure Industrial Control Systems (ICS); the SWaT testbed is a pivotal component within a broader research initiative conducted at

iTrust [25]. This initiative is dedicated to advancing the design of secure cyber-physical systems.

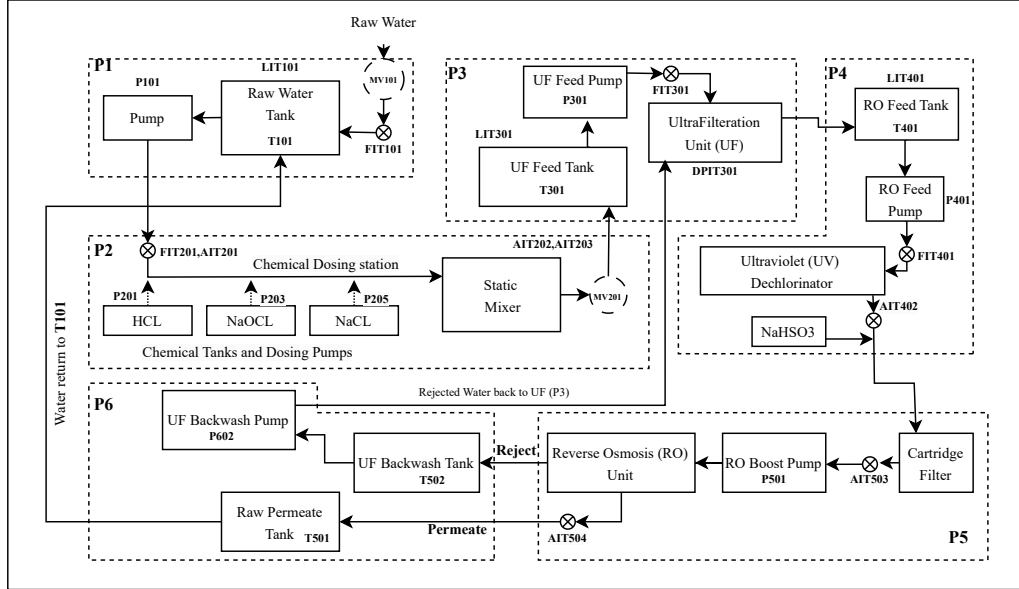


Figure 2. **SWaT Process Diagram.** P1-P6 indicate the six stages in the treatment process. Solid arrows indicate flow of water or chemicals in the dosing station. Dashed arrows indicate potential cyber attack points. **LIT**: Level Indicator and Transmitter; **Pxxx**: Pump; **AITxxx**: Property indicator and Transmitter; **FITxxx**: flow meter; **DPIT**: Differential Pressure Indicator and Transmitter.

Water Treatment Process: As shown in the figure 2, SWaT Comprises six stages labeled P1 through P6; each stage is managed by its dedicated set of Programmable Logic Controllers (PLCs). Stage P1 is responsible for regulating the inflow of water into the treatment process by manipulating a valve (not depicted) that connects the inlet pipe to the raw water tank. Subsequently, water from the raw water tank undergoes pumping through a chemical dosing station (stage P2, chlorination) to reach another Ultra Filtration (UF) Feed water tank situated in stage P3. Within this stage, a UF feed pump transports water through UF membranes to the Reverse Osmosis (RO) feed water tank in stage P4. At this juncture, an RO feed pump propels the water through an ultraviolet de-chlorination unit under the control of a Programmable Logic Controller (PLC) in stage P4. This step is crucial for eliminating free chlorine from the water before it traverses the reverse osmosis unit in stage P5. In stage P4, Sodium bisulfate (NaHSO₃) may be introduced to regulate the Oxidation Reduction Potential (ORP).

Within stage P5, the de-chlorinated water undergoes filtration through a 3-stage RO

unit. The filtered water from the RO unit is stored in the permeate tank, while the reject is directed to the UF backwash tank. Stage P6 oversees the cleaning of membranes in the UF unit, controlling the UF backwash pump's activation and deactivation. The backwash cycle is automatically initiated every 30 minutes, taking less than a minute to complete. Differential pressure sensors in stage P3 measure the pressure drop across the UF unit. A backwash cycle is also initiated if the pressure drop exceeds 0.4 bar, indicating that the membranes need immediate cleaning. A differential pressure meter installed in stage P3 is used by PLC-3 to obtain the pressure drop [25].

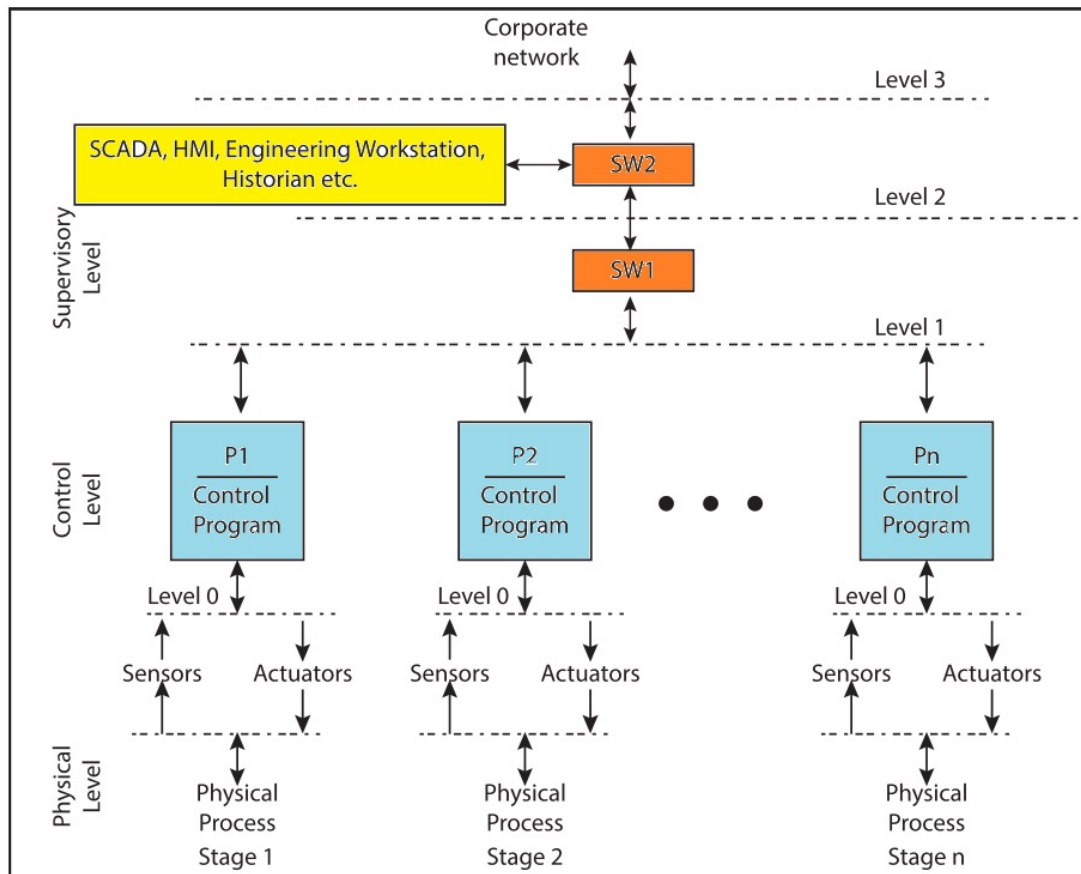


Figure 3. Control Portion Architecture of a CPS. P1, P2, . . . , Pn denote PLCs. SW1 and SW2 are Switches. Each PLC communicates with its sensors and actuators through a local network at Level 0. PLCs communicate among themselves via another network at Level 1. Communication with SCADA and other computers is via a Level 3 network not shown here. Note that the actuators, e.g., a pump, also have sensors to indicate their condition, adapted from [8].

Communications: The control system functions as an assembly of Programmable

Logic Controllers (PLCs), each assigned to oversee a specific segment of the Cyber-Physical System (CPS), as depicted in Figure 3. In this configuration, every PLC establishes communication with a set of sensors and actuators through a localized network, denoted as Level 0 or the field-bus network. Additionally, the PLCs engage in communication with one another through the Level 1 network. This hierarchical network structure aligns with established norms in industrial control systems, reflecting a layered approach [8].

As in figure 3, Every Programmable Logic Controller (PLC) is responsible for acquiring data from sensors affiliated with its respective stage while simultaneously regulating pumps and valves within its domain. Ultrasonic level sensors installed in each tank relay information about the water level to the corresponding PLCs [8]. Furthermore, various sensors are deployed to assess water's physical and chemical attributes traversing the six stages. PLCs engage in intercommunication via a dedicated network, and the exchange of information between sensors, actuators, and PLCs can occur through either wired or wireless connections. The system seamlessly transitions between wired and wireless modes through manual switches.

2.3 Deception

Deception technology functions as a strategic approach aimed at diverting cyber adversaries from an organization's genuine assets towards decoys or traps. These decoys emulate authentic servers, applications, and data, leading intruders to believe they have breached critical assets when, in fact, they have not. This tactic is implemented to mitigate harm and safeguard the organization's actual assets [24]. Compared to conventional security measures, deception techniques operate with subtlety and complement traditional security protocols. Typically, they entail the deployment of decoys or honeypots replicating network infrastructure or services supplemented with fabricated data. Cyber deception serves multiple purposes. It not only redirects attackers away from authentic data but also instills confusion in their endeavors, impeding the progress of their attacks [24]. Consequently, a convoluted environment is established, compelling threat actors to expend resources on inconsequential targets.

In addition to deterring adversaries, the deployment of decoys facilitates the monitoring of attacker behavior. This enables security teams to gain insights into adversaries' tactics, techniques, and procedures, enhancing the organization's defensive capabilities [24]. Functioning as a facet of threat detection and intelligence gathering, cyber deception technology is particularly potent in revealing the psychological dynamics of attackers and acquiring real-time threat intelligence from their activities. Figure 4, published in article [27], explains how the organization's network looks before and after the deployment of the deception.

Cyber deception technology fundamentally alters the dynamics of cyber attacks by imposing additional resource requirements on attackers. This strategic approach

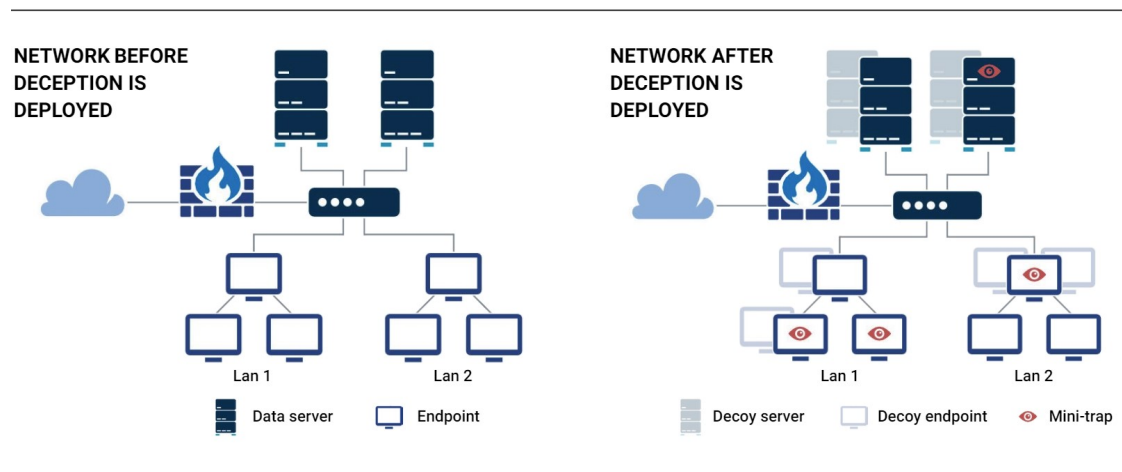


Figure 4. Example architecture for before and after deployment of deception, adapted from [27].

compels threat actors to expend time and effort, thereby acting as a deterrent in its own right. Moreover, it diminishes the attractiveness of an organization as a target while simultaneously enhancing the protection of legitimate data, rendering it more challenging for adversaries to locate and compromise [27]. An integral aspect of cyber deception is its ability to establish decoys at entry points, fortifying the organization’s attack surface. This proactive measure serves as a supplementary defensive layer alongside existing security tools. In the event of a breach, deception technology assumes a defensive stance, impeding adversaries’ progress throughout the organization’s network [24]. Notably, it represents the most effective defense against insider threats.

In cybersecurity, deception technologies offer unparalleled detection capabilities. They not only furnish valuable insights into adversaries’ tactics and behaviors but also exhibit a notable attribute: they exclusively respond to genuine malicious activities. Consequently, they mitigate the burden of alert fatigue experienced by security teams, arising from the influx of false positive alerts generated by various monitoring tools and vulnerability scans [24].

Deception techniques are most effective when employed in tandem, as their combined usage introduces complexity into the security network, thereby confounding adversaries, even those familiar with deceptive tactics. By implementing multiple types of deception, organizations can introduce a myriad of variables, all while monitoring adversaries’ interactions with the decoys [24].

Here are the principal security tools driving cyber deception initiatives today:

- **Honeypots:** Honeypots are strategically placed traps designed to identify, divert, or counteract attempts to compromise data or exploit information systems. These decoys typically present data that appears valuable to attackers but is actually

isolated and under surveillance by defenders. By enticing attackers to engage with the honeypots, defenders gain valuable insights into their methodologies and may preemptively thwart impending attacks [24].

- **Honeynets:** Honeynets consist of interconnected honeypots deployed collectively to divert attackers away from critical data and systems, amplifying the effectiveness of deception strategies [27].
- **Masking:** Masking involves concealing genuine assets or data that require protection. By rendering authentic data invisible within the network, this technique removes it from immediate detection without raising suspicion, constituting the initial step toward effective deception [27].
- **Mimicking:** Mimicking entails replacing concealed assets with decoys that closely resemble genuine components of the network, thereby maintaining an attractive attack surface for adversaries under surveillance. The efficacy of mimicking lies in its ability to render false assets indistinguishable from authentic ones [24].
- **Inventing:** Inventing involves fabricating entirely new assets that emulate the appearance of genuine components. These fictitious assets, such as simulated services, serve to divert attention from actual entry points while maintaining an enticing facade for adversaries [24].
- **Repackaging:** Repackaging involves modifying genuine assets to obscure their true value, thereby rendering them inconspicuous to adversaries. Assets that cannot be effectively masked may be repackaged to diminish their apparent significance, facilitating their oversight by intruders [27].
- **Dazzling:** Dazzling represents the least subtle form of deception, inundating attackers with overwhelming information to obscure genuine assets amidst a sea of falsified data. While less discreet, dazzling remains effective against less sophisticated threats, overwhelming adversaries with an information overload that obscures genuine targets [24].

2.4 Digital Twin

A Digital Twin is a digital or virtual representation of a physical object, process, service, or environment that behaves and looks like its counterpart in the real world. A DT is a computer program that harnesses real-world data to create simulations that can accurately predict how a product or process will perform. This simulation relies on current asset conditions and historical data, often integrating technologies like the Internet of Things (IoT), artificial intelligence, and software analytics to enhance its capabilities. DTs are

powerful tools for innovation, enabling engineers to understand and improve the performance of real-world systems by simulating and monitoring their virtual counterparts [30].

2.4.1 Architecture and Components

In Grieves' seminal white paper [16], the initial architecture of DTs was introduced, conceptualized as a three-dimensional model comprising physical space, virtual space, and a connecting interface facilitating data harmonization between these domains (refer to Figure 5). Subsequently, Tao et al. extended this foundational model to encompass five dimensions, incorporating physical space, virtual space, connection, DT data, and service elements (as depicted in Figure 6). The tangible, or physical, layer contributes real-world asset data to the virtual counterpart, while the virtual layer aids in replication and decision-making processes for the physical domain. The service component supports these layers, which facilitates their operation, evolution, and optimization over time. Central to the architecture is the DT data repository, serving as the primary source of internal information, while the connection element interconnects all model components, establishing the desired interconnected loop [32]. For contemporary complex systems, essential components include sensors for real-world data acquisition, a physical twin, edge processing capabilities, data security measures, the DT itself, data storage and processing infrastructure, and interfaces for reporting. Additionally, effective visualization mechanisms are integral to the user experience. The communication element serves as the nexus where the physical and virtual layers converge, offering various protocols and interfaces such as Wi-Fi, Bluetooth, and wired connections.

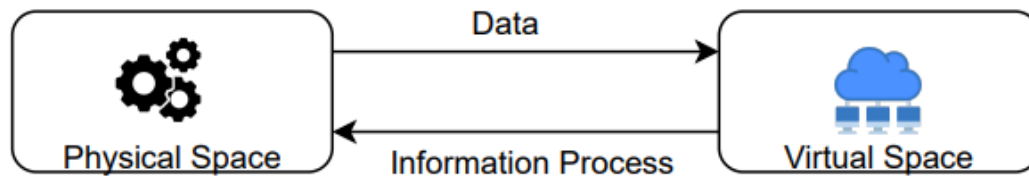


Figure 5. 3D DT Model Representation, adapted from [16]

There can be different ways to leverage DT technology to enhance the cybersecurity of water systems. While DTs are typically used for simulation and predictive purposes, they can also serve as invaluable tools for monitoring and diverting cyberattacks[30].

- **Predictive Simulation:** Develop a DT of the water supply system to simulate its operations and identify potential vulnerabilities or attack scenarios. This simulation can help us predict how cyberattacks might impact the physical system, which can help us to test various security measures virtually[30].

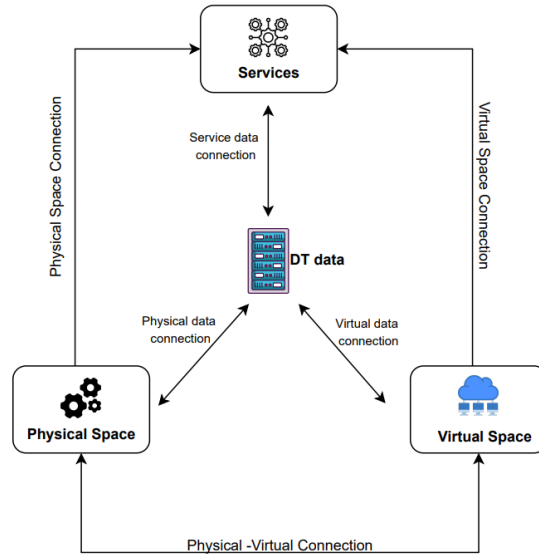


Figure 6. 5D DT Model Representation, adapted from [32]

- **Deception decoy Implementation:** We can use the DT as an attack deception decoy, essentially a decoy system that mimics the real system. It can attract potential cyber attackers and divert their attention away from the physical water supply infrastructure. We can gather valuable insights into threat actors' tactics and intentions by monitoring this DT for attack attempts.
- **Real-time Monitoring:** We can Integrate IoT sensors into the DT to capture real-world data. This data can be used for continuous monitoring and analysis to detect anomalies and potential security breaches. This provides a way to respond to threats in real time [30].
- **Security Testing:** Implementing the DT to test and assess various security measures in a controlled environment, such as intrusion detection systems and access controls. This leads us to evaluate the effectiveness of these measures before implementing them in the physical system.

Using a DT as both a predictive and protective tool, there are several applications to enhance the cybersecurity of water systems. This work focuses on how the DTs can be implemented as an attack deception decoy.

2.4.2 Digital Twin Advantages Over Traditional Honeypot

This section sheds light on the advantages of using the DT as an attack deception mechanism over traditional Honeypots. In the table 1, the various advantages of a DT

over a traditional honeypot are mentioned, which can be explained in depth as follows.

Table 1. Advantages of Digital Twins over traditional Honeypots as Attack Deception Technique

Advantages Aspect	Honeypots	Digital Twins
Realistic Simulation	May lack the complexity of real systems.	Replicates actual systems realistically.
Behavioral Modeling	Often has static or predefined behaviors.	Allows Dynamic Behavioral Modeling.
Integration with Operational Systems	Typically standalone entities.	Can be seamlessly integrated.
Data Analysis and Forensics	May generate limited data for analysis.	Generates detailed data for analysis.
Reduced False Positives	May trigger false positives due to their static nature.	Aims to minimize false positives by replicating authentic environments.
Adaptability and Scalability	May require significant effort to adapt.	Easily adaptable to various scenarios.

- **Realistic Simulation:** DTs replicate actual systems and processes, providing a highly realistic environment for attackers. This authenticity increases the chances of attracting and detecting sophisticated attacks. On the other hand, Honeypots, while useful, may lack the complexity and authenticity of real systems. Skilled attackers can recognize honeypots and adjust their behavior accordingly.
- **Behavioral Modeling:** DTs allow for dynamic behavioral modeling, mimicking the genuine behavior of systems and applications. This enables more accurate detection of anomalies and malicious activities. Whereas, Honeypots often have static or predefined behaviors, making them more predictable to attackers who may recognize and avoid them.
- **Integration with Operational Systems:** DTs can be integrated seamlessly with operational systems, providing a comprehensive view of both normal and deceptive activities. This integration enhances the overall security posture. And while we talk about Honeypots are typically standalone entities, and integrating them with operational systems can be challenging. This separation may limit their effectiveness in a holistic security strategy.
- **Data Analysis and Forensics:**DTs generate detailed data on attacker interactions, facilitating in-depth analysis and forensic investigations. This information can be

valuable for understanding attack patterns and improving overall security. While Honeypots may generate limited data, and the lack of comprehensive information can hinder thorough analysis and forensic efforts.

- **Reduced False Positives:**DTs, by replicating authentic environments, aim to minimize false positives. This ensures that alerts raised are more likely to indicate genuine malicious activities. Whereas, Honeypots may trigger false positives due to their static nature or because attackers recognize and avoid them, leading to alerts that may not necessarily represent actual threats.
- **Adaptability and Scalability:**DTs can be easily adapted to represent various systems and scaled to simulate diverse environments, making them versatile for different use cases and scalable for larger networks. On the other hand, Honeypots may require significant effort to adapt to different scenarios, and their scalability might be limited in complex network architectures.

In summary, DTs provide a more advanced and adaptable approach to attack deception, leveraging realistic simulations and dynamic modeling to enhance security measures.

2.5 Security Operations Center (SOC) Playbook

This section explains the Introduction to Security Operation Center (SOC) and playbook with its purpose, scope, and how it can be utilized. Security Operation Center (SOC) is a centralized function within an organization employing people, processes, and technology to continuously monitor and improve an organization's security posture while preventing, detecting, analyzing, and responding to cybersecurity incidents [38]. A SOC acts like the hub or central command post, taking in telemetry from across an organization's IT infrastructure, including its networks, devices, appliances, and information stores, wherever those assets reside [38].

The SOC holds a pivotal position in the realm of cybersecurity, acting as a central hub for monitoring, detecting, and responding to security incidents. Within the arsenal of a SOC, the SOC playbook emerges as a fundamental instrument. Crafted meticulously, the SOC playbook serves as a comprehensive guide, meticulously detailing the sequential procedures essential for security analysts to adeptly navigate and mitigate security incidents [39]. This structured document delineates the precise actions to undertake, the requisite tools to deploy, and the pertinent personnel to engage during the course of an incident. The SOC playbook serves as a strategic road map, furnishing the SOC with the requisite directives to effectively counter a myriad of cyber threats.

The purpose of the SOC playbook is to provide a standardized and structured approach for security analysts within a SOC to effectively respond to cybersecurity incidents. By delineating step-by-step procedures, tools, and personnel responsibilities, the playbook

aims to streamline incident response processes, minimize response times, and ensure consistent and coordinated actions during security incidents. Additionally, the playbook serves as a reference guide, enabling security analysts to leverage best practices and lessons learned from previous incidents to enhance incident response capabilities over time.

The scope of the SOC playbook encompasses various aspects of incident response, including but not limited to incident detection, analysis, containment, eradication, and recovery. It is designed to cater to a wide range of cybersecurity incidents, spanning from low-level security breaches to sophisticated cyberattacks. Moreover, the playbook is adaptable to accommodate the evolving threat landscape and organizational requirements, allowing for continuous refinement and optimization of incident response strategies. Ultimately, the SOC playbook serves as a crucial tool for enhancing the overall security posture of an organization by facilitating swift and effective responses to cybersecurity incidents.

2.6 Blockchain

Blockchain is a decentralized and distributed ledger technology that ensures secure and transparent record-keeping. Blockchain, initially proposed by Satoshi Nakamoto in 2008 [26], integrates various existing technologies such as distributed ledgers, cryptography, hashing, and consensus protocols. In this architecture, transaction records are organized into blocks and distributed across a peer-to-peer network, with each node holding a copy of the entire chain. To ensure the integrity of the ledger, transactions undergo validation through a consensus mechanism, which comprises a predefined set of rules and policies [9]. The consensus mechanism, agreed upon by stakeholders before deployment, varies across different blockchain technologies [29]. Once validated by the majority of participating nodes, a transaction is added to the blockchain as a new block, complete with a timestamp, hash of the previous block, and transaction data. This process ensures the creation of a secure, decentralized, persistent, fault-tolerant, and auditable ledger, enabling decentralized automated transactions without the need for centralized control [9]. Its core features of decentralization, immutability, and transparency can enhance the cybersecurity aspects of DT systems. Key characteristics of a fundamental blockchain can be delineated as follows:

- **Decentralization:** In contrast to traditional transactions reliant on a central authority for validation, blockchain technology engages multiple distributed nodes to authenticate each transaction, ensuring fault tolerance and transparency, along with user control flexibility and resilience against attacks [22].
- **Distributed:** By generating a distributed record that is owned and verified by other users, blockchain eliminates the need for third parties in transactional processes [22].

- **Persistence:** Through consensus mechanisms, timestamps, and cryptographic seals, blockchain ensures the creation of immutable transaction blocks, fostering data persistence, fault protection, and authorized transaction ownership [29].
- **Pseudo-anonymity:** Transactions and validations within the blockchain maintain stakeholder anonymity, employing hash addresses to safeguard user information confidentiality, thus distinguishing blockchain from conventional transaction systems [22].
- **Traceability:** Each block in the chain contains references to the preceding block, enabling sequential storage of transactions and facilitating the traceability of tampering or malicious activity within the blockchain [22].

Furthermore, owing to its secure distributed framework facilitating information sharing and coordination among participating entities, blockchain technology has garnered significant attention from scholars and professionals across various disciplines, including finance, law, and computer science. The rapid adoption of blockchain can largely be attributed to the considerable success of its prototype, Blockchain, the precursor to cryptocurrencies. The emergence of alternative platforms such as Ethereum (ETH) and Ripple has broadened the adoption of blockchain technology beyond the realm of finance, extending its utility to non-financial domains such as intellectual property, proof of location, voting, and healthcare.

2.6.1 Leveraging the Data Immutability feature of Blockchain

Blockchain technology offers an approach to ensuring data immutability, a fundamental property whereby once data is recorded on the blockchain, it becomes unalterable. This inherent feature of blockchain not only enhances data security but also establishes trust in the integrity of recorded information. In the realm of water systems, blockchain's immutability finds significant application, particularly in recording data from physical sensors measuring water quality and system parameters. By storing this data as transactions on the blockchain, a tamper-proof record is created, safeguarding historical information vital for analysis, compliance, and legal purposes.

Within the context of DTs for water systems, blockchain's immutability guarantees that simulations, predictions, and analyses are grounded on reliable and unaltered data. Integrating blockchain into the DT ecosystem of water systems introduces an additional layer of security, reinforcing the integrity and trustworthiness of data essential for optimizing system operations and addressing cybersecurity challenges. For instance, imagine a scenario where water quality data from sensors is recorded on a blockchain. Attempts to tamper with this data would be immediately detected due to the blockchain's immutable nature, ensuring the reliability of information used within the DT for decision-making and system optimization.

2.6.2 Types Of Blockchain

Based on their characteristics and governance policies, blockchain technology is typically categorized into three main types:

- **Public:** Public blockchains are characterized by their openness, allowing unrestricted participation from any individual to join the decentralized ledger network. These blockchains operate on a permissionless basis, where no limitations are imposed on access or validation of transactions. Public blockchains maintain features such as immutability, transparency, and security through distributed consensus algorithms. Examples of public blockchains include Bitcoin (BTC) and Ethereum (ETH).
- **Private:** Also known as permissioned blockchains, these networks are restricted environments where participants require validation and authorization to access the network. Permissioned blockchains are commonly utilized within organizations where strict control over network access is essential. They are often employed for internal purposes such as supply-chain management or intra-organizational collaborations. Examples of private blockchain frameworks include Hyperledger Fabric and R3 Corda.
- **Hybrid:** Hybrid blockchains, also known as Consortium blockchains, amalgamate features from both public and private blockchains. In consortium blockchains, multiple organizations or entities form a consortium and jointly govern the network. This model is particularly beneficial for large organizations with diverse stakeholders, allowing for shared control and privacy levels. Consortium blockchains typically exhibit improved transaction throughput and scalability compared to public and private blockchains, while maintaining a certain degree of decentralization.

2.6.3 Ethereum

Ethereum is a public blockchain that extends beyond its original application in cryptocurrency and payments, offering an open-source programmable platform for various digital assets. Its decentralized nature enables transactions to occur without the need for intermediaries. The building components of Ethereum include:

- **Ether (ETH):** The native currency of the Ethereum platform and it is denoted as ETH.
- **Gas:** A fixed fee associated with each transaction, determined by the current market value of Ether, serves to safeguard the network against potential attacks like distributed denial of service (DDoS).

- **Transactions:** Formal agreements between parties to exchange digital assets.
- **Ethereum Accounts:**
 - **Externally Owned Accounts (EOA):** Accounts controlled by individuals holding Ether balances, each associated with a unique private/public key pair.
 - **Contract Accounts (CA):** Similar to EOAs but devoid of private and public keys.
- **Ethereum Nodes:**
 - **NVM:** Responsible for executing functions embedded within smart contracts, with limited network access restricted to Externally Owned Accounts (EOA), Contract Accounts (CA), and their respective storage.
 - **Mining Nodes:** Participants with direct access to the blockchain, rewarded with gas for validating transactions through the solution of cryptographic puzzles.
- **Blocks:** Sequential collections of transactions forming the Ethereum blockchain, with the initial block termed the genesis block.
- **Smart Contract:** Digital representations of business rules encoded in the Solidity programming language, requiring consensus among account holders to execute transactions autonomously.

2.6.4 Hyperledger Fabric

Hyperledger Fabric (HLF) stands out as a permissioned blockchain solution, characterized by its distributed architecture and robust access control mechanisms, which prevent unauthorized access to the network. Its architecture, featuring channels and organizations, facilitates independent transactions within distinct organizational boundaries without disrupting overall network operations. Written in languages such as Go, Node.js, and Java, HLF offers customization options and automation capabilities for various business processes. Notably, its unique execute-order-validate algorithm enhances transaction efficiency and reliability. The key components³ of a Hyperledger Fabric network include:

- **Assets:** Represent valuable entities within a business context, stored as stateful key-value pairs in the ledger.
- **Shared Ledger:** Consists of the World State and Blockchain, housing transaction records and asset states.
- **Smart Contract:** Known as chaincode, it encapsulates the business logic and facilitates interactions with the ledger to execute transactions.

³<https://developer.ibm.com/articles/blockchain-basics-hyperledger-fabric/>

- **Peer Nodes:** Core components hosting chaincode and ledger data, comprising endorsing, committing, and ordering peer nodes, each serving distinct roles in transaction processing.
- **Channel:** A logical grouping of peer nodes enabling secure and private transactions within a subset of the network.
- **Organizations:** Network members with one or more peer nodes, facilitating inter-organizational transactions and collaborations.
- **Membership Service Provider (MSP):** Manages user authentication and enrollment, ensuring secure connections between network components and users.
- **Ordering Service:** Responsible for transaction ordering and dissemination across the network, leveraging mechanisms like Solo and Kafka for configuration and management.

These components collectively contribute to the functionality and integrity of the Hyperledger Fabric blockchain network, supporting diverse enterprise use cases with enhanced security and scalability.

2.6.5 Why Ethereum?

The selection of Ethereum as the blockchain technology for storing critical property values within the water CPS system stems from its robustness and suitability for smart contract deployment. Ethereum's established platform facilitates the secure encoding of predefined component properties, ensuring immutability and integrity. Moreover, its public blockchain architecture offers transparency and decentralization, aligning with the system's goal of preventing unauthorized tampering.

Furthermore, Ethereum's support for programmable transactions and its native cryptocurrency, Ether, enables dynamic threshold values and triggers for detecting anomalies in component properties. This enhances the system's ability to respond to security breaches efficiently. With an active developer community and extensive documentation, Ethereum provides accessibility and support for integration into complex cyber-physical systems like water CPS. Leveraging Ethereum's capabilities fortifies critical infrastructure against cyber threats while ensuring secure and reliable operation.

3 Systematic Literature Review

In today's world, marked by unprecedented technological integration, the seamless interaction between digital processes and physical components such as CPS is pivotal in ensuring the uninterrupted flow of clean and safe water to communities worldwide.

However, this intersection of the digital and physical realms also exposes these systems to an escalating threat landscape of cyberattacks, as seen in recent days. This section serves as a crucial exploration of existing knowledge and research in the domain of cybersecurity for water systems, with a specific focus on CPS, and plays a pivotal role in laying the foundation for the overall work, providing an in-depth understanding of the historical landscape and existing strategies proposed in safeguarding water supply infrastructures. Three fundamental research questions guide this investigation: RQ1.1: What are the primary security threats to water systems? RQ1.2: What vulnerabilities in water systems are exploited by threat actors? RQ1.3: What are the detection mechanisms proposed for water systems against security threats? We follow the Kitchenham review guidelines to perform SLR [21]. By extensively reviewing the literature, this section contributes a synthesized understanding of historical cyber threats and defense strategies in the context of water systems. The results obtained from this exploration will inform the subsequent phases of this work, guiding the development of robust cybersecurity measures and strategies using DT technology.

3.1 Review Questions

We have developed certain Research Questions for the SLR, which are defined as follows:

RQ1.1: What are the primary security threats to water systems?

RQ1.2: What vulnerabilities in water systems are exploited by threat actors?

RQ1.3: What detection mechanisms are proposed for water systems against security threats?

This work addresses three key research questions regarding the cybersecurity of water systems. Firstly, [RQ1.1] seeks to identify and classify the primary security threats that target water systems, providing a comprehensive understanding of the risks these critical infrastructures face. Secondly, [RQ1.2] the study focuses on identifying specific vulnerabilities within water systems that are exploited by threat actors, aiming to pinpoint weaknesses and areas of potential compromise. Lastly, [RQ1.3] explores existing detection mechanisms proposed for mitigating security threats to water systems, assessing their effectiveness in detecting and responding to cyberattacks on water infrastructure. Through these inquiries, the study aims to contribute to developing robust cybersecurity strategies tailored to the unique challenges water systems face.

3.2 Review Settings

This section encapsulates the methodological framework guiding this work's systematic exploration and analysis of pertinent literature. Comprising several key subsections, this section delineates the essential steps undertaken to ensure a thorough and methodical review process. Firstly, the search strings subsection defines the specific terms and logical operators employed to query academic databases, establishing the scope of the search

and guiding the retrieval of relevant literature. Secondly, the subsection on data sources identifies the primary repositories and databases utilized to access scholarly literature, ensuring a comprehensive exploration of the research topic.

Within this methodological framework, the inclusion and exclusion criteria subsection outlines the parameters used to assess the eligibility of retrieved papers for inclusion in the review, refining the selection process to include only literature meeting predefined standards. Subsequently, the paper selection subsection details the systematic approach employed to sift through retrieved papers and identify those meeting the inclusion criteria. Finally, the data extraction strategy subsection describes the systematic process of extracting relevant information from selected papers for analysis, ensuring structured data collection to facilitate subsequent synthesis and interpretation. Through these cohesive subsections, this section provides a transparent and rigorous methodology for systematically exploring and synthesizing literature pertinent to this work.

3.2.1 Search Strings

In the search for relevant literature, the following search string was employed: (("water supply system" OR "water infrastructure" OR "Water") AND ("cybersecurity challenges" OR "security threats" OR "cybersecurity" OR "attack" OR "Vulnerability" OR "Security") AND ("digital twin" OR "cyber-physical system" OR "DT" OR "CPS")). This string comprises three main components enclosed within parentheses and combined using logical operators. Synonymous terms within each component, such as "water supply system," "water infrastructure," and "Water," were included to capture variations in terminology. Similarly, a range of terms related to cybersecurity, including "cybersecurity challenges," "security threats," and "attack," among others, were incorporated. Additionally, terms associated with DT technology and cyber-physical systems, such as "digital twin," "cyber-physical system," and "DT," were included to focus the search on relevant topics. Papers meeting the criteria were initially selected based on the presence of these terms in the title or abstract, ensuring relevance to the study's objectives.

3.2.2 Data Sources

The preliminary exploration for relevant academic papers involved utilizing the IEEE digital library, ScienceDirect, and ACM digital library. The main purpose in selecting these three databases stems from their status as prominent sources for recent state-of-the-art technological papers. Supplementary relevant literature was chosen by examining the related work sections and citations within the initially identified papers. Since the concept of integrating DT with Cyber-security in the water sector and CPS is a relatively new paradigm, Additionally, gray literature was considered, meticulously ensuring that the outcomes aligned with the specified inclusion criteria.

3.2.3 Inclusion and Exclusion Criteria

In conducting the review, the scope encompassed recent literature spanning the years 2017 to 2024, with a particular emphasis on exploring the domain of cybersecurity in the water sector CPS and the role of DTs in water cybersecurity. The primary objective was to identify scholarly works explicitly referencing cybersecurity measures and their efficacy in safeguarding CPS. Notably, the review excluded early access publications and book sections to ensure a focus on rigorously reviewed research that aligns closely with the established inclusion criteria. This approach aimed to yield comprehensive insights into the role of cybersecurity in enhancing security within various contexts, particularly in the realm of the water sector, thereby contributing to a nuanced understanding of contemporary advancements in the field.

Inclusion Criteria

- IC1: Literature related to cybersecurity in the water sector.
- IC2: Literature addressing the application of digital twins in enhancing security measures for water infrastructure.
- IC3: Literature exploring the integration of digital twins and cyber-physical systems (CPS) within water infrastructure.

Exclusion Criteria

- EC1: Literary items published before 2017 (436)
- EC2: Literature that is not journals, magazines, or conference papers (316)
- EC3: Literature without explicit mention of Cybersecurity of CPS in Water sector (124)
- EC4: Literature that is tagged Early Access (78)

3.2.4 Paper Selection

The initial search across digital libraries yielded numerous results, but not all were relevant. After applying the exclusion criteria (EC1), I was left with 436 papers. Subsequently, EC2 filtered conference materials, magazines, and academic journals, resulting in 316 papers. Further refinement through EC3 and EC4 narrowed the selection to 78 relevant papers. Each of these underwent manual evaluation based on predefined inclusion criteria, primarily analyzing the introduction and abstracts. Ultimately, 12

papers met the criteria and were retained for further study. Additionally, snowballing and reference scrutiny led to the inclusion of 7 more relevant papers within the review's scope.

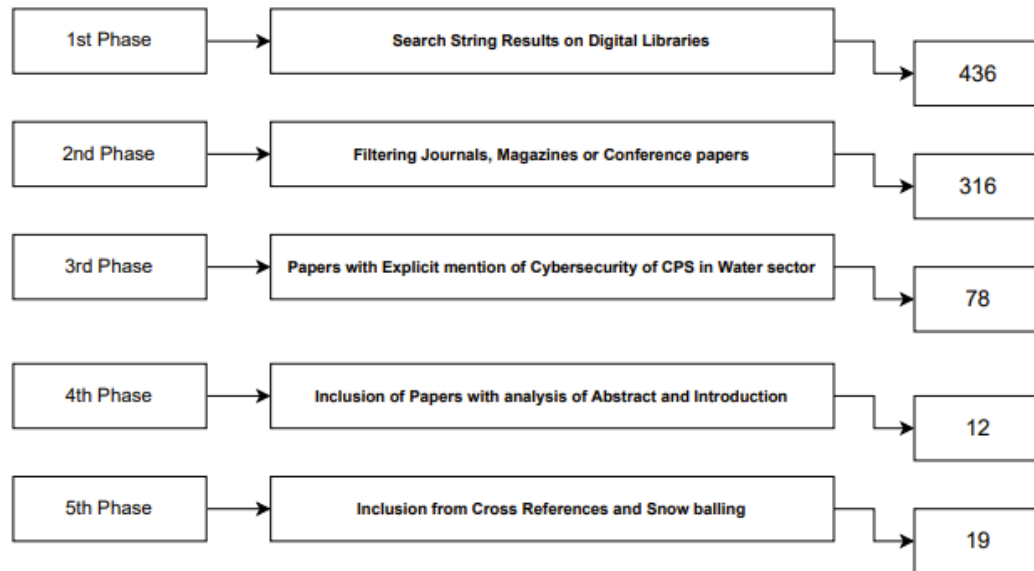


Figure 7. Paper Selection Process

3.2.5 Data Extraction Strategy

Nineteen papers were read through to gather data for research questions. The table was formed for further analysis to note the data collected from papers that could answer the review questions developed. The data extraction table (as Table 2 shows) consists of the data items as Research Work, Application, Objective, Contribution, Threats targeting water systems, mainly CPS and Threat detection and mitigation strategies defined. These columns were chosen in accordance with the RQs.

3.3 Presentation of Results

In this section, the outcomes of the literature review are presented in the form of tables and descriptions. The tables contain data aiming at providing the answers for the RQs depicted in the section Review Questions. The main agenda of the literature review was to find the existing research where past cyber attacks on water systems are categorized and different detection mechanisms or strategies proposed to defend them.

Firstly we categorize the papers according to the purpose and scope of study, Relating to the Review Questions developed specifically for the Literature Review studies.

Table 2. Data Extraction Form

Data Item	Value
Research Work, Year	Paper and Year of Publication
Application	Area of Research
Objective and contribution	Objective and contribution of the literature
Threats targeting water systems	Threats discussed in the literature
Threat detection and mitigation strategies defined	The main threat detection and mitigation strategy proposed to defend water systems against cyber threats

3.3.1 Water systems: Challenges and Operations

The escalating frequency of reported attacks directed at critical cyber-physical systems essential for national infrastructure services underscores a growing concern. Evidenced by impactful attacks like Stuxnet [12], DuQu [33], BlackEnergy [20], and Havex [11], these incidents highlight the potential for catastrophic consequences.

Firstly the table 3 was created to provide information regarding the cyber attacks that occurred in the past targeting water systems by categorizing them according to the location and year, Facility Targeted, the component of water supply system targeted, Domain and attacker action, Details of the Incident, Impact of the Incident and Reason behind the attack was successful which can also be considered as Vulnerability that is either physical or technical.

RQ1.1: What are the primary security threats to water systems? and RQ1.2: What vulnerabilities in water systems are exploited by threat actors?

Publicly disclosed incidents targeting water infrastructure services, as detailed in Table 3 and Table 4, underscore the potential success of attacks exploiting a range of vulnerabilities. These attacks pose a dual threat by directly disrupting services and causing harm to control equipment and communication networks, subsequently impacting critical services. The ramifications extend beyond the immediate disruptions to encompass public health and environmental well-being and entail significant financial and reputational losses for the affected companies. The Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) in the United States is a notable resource for documenting cybersecurity incidents in the water sector. This repository serves as a widely acknowledged reference for understanding the landscape of cyber threats in the water industry.

Table 3. Past Cyber-attacks on Water Systems

*	Location and Year	Name and Type	Details
1	Australia 2000	Maroochy Water Services(Wastewater Treatment facility)	3rd party contractor, installed PDS Compact 500 RTUs at all 142 sewage pumping stations. This enabled to remotely control and monitor the pumps through a SCADA system.
2	PA, U.S. 2006	Pennsylvania Water Filtering Plant(Water treatment)	Hackers planted a computer virus on the laptop of an employee and then installed a malicious software on the plants computer system.
3	CA, U.S. 2007	Tehama-Colusa Canal(Irrigation System)	The TCAA employee accessed the computer system and installed unauthorized software on the SCADA system.
4	FL, U.S. 2012	Key Largo Wastewater Treatment District(Wastewater treatment)	Stolen login credentials were used to access district's computer system.
5	NY, U.S. 2013	Bowman Avenue Dam	hackers obtained unauthorized remote access to the SCADA system; a cyber-attack that allowed them to gather information on water levels, temperature, and the status of the sluice gate.

*	Location and Year	Name and Type	Details
6	U.S 2014	Five water utilities(Water utility)	The attack was caused by a fired employee of the company that manufactured the smart water meters. using his access to the base station network, he conducted various malicious activities, such as changing the root passwords, modifying the TGB radio frequency, and overwriting computer scripts.
7	U.S 2016	Kemuri Water Company(Water utility)	Possible unauthorized access to the systems as well as a series of unexplained valve manipulation patterns caused by several high-risk vulnerabilities on an outdated mid-range computer system (AS400).
8	U.S 2016	An undisclosed utility(Water utility)	The administrator found heavy network traffic originating from the control panel of a pumping station.

*	Location and Year	Name and Type	Details
9	U.S 2016	An undisclosed drinking water utility(Water utility)	Water Utility noticed a 15,000 percent increase in their monthly cellular data bills. The authority was hacked between November 2016 and January 2017. The utility had seven Sixnet BT series cellular routers, which provided wireless access for monitoring the utility's pumping stations as well as a few other sites. Four of these seven routers were compromised by the hackers
10	Uk 2017	A regional water supplier(Water supplier)	Bank details of the Water supplier employees were changed by the CRM partner employee.
11	Europe 2018	A European water utility(Water utility)	Cryptocurrency Malware (CoinMiner) was installed on HMI on the SCADA Network
12	NC, U.S. 2018	Onslow Water and Sewer Authority(Water utility)	A sophisticated ransomware attack which locked out employees and encrypted databases, leaving the utility with limited computing capabilities. The hack began with persistent cyber-attacks through a virus known as EMOTET.

*	Location and Year	Name and Type	Details
13	CO, U.S. 2019	Fort Collins Loveland Water District(Water District)	The utility had fallen victim to a ransomware cyber-attack. The hackers demanded a ransom to restore access.
14	FL, U.S. 2019	Riviera Beach Water Utility(Water utility)	An employee of the police department opened an infected email. Paralyzing computer systems of the police department, city council, and other local government offices, the ransomware sent all operations offline and encrypted their data. The attack also spread to the water utility.
15	Israel 2020	Pennsylvania Water Filtering Plant(Water Utility)	Israel government reported cyber-attacks against water supply and treatment facilities and urged these facilities to change passwords.

Table 4. Past Incidents Analysis

*	*Target @Domain # Attacker Action	Impact	Attack success Reason
1	* RTU/PLC @ OT # Configuration Change	Environmental pollution. Nearly one million liters of raw sewage into the river, local parks, and residential grounds. 500 meters of open drain in a residential area was polluted.	No cybersecurity procedures, policies, or defenses were present.
2	* Workstations @ IT # Data Exfiltration	Data breach	Entry point to the plant's computer system was an employee's laptop. It is considered a weak link in the security chain.
3	* SCADA @ OT # Software Installation	Water theft. The electrical supervisor at the TCCA accessed and damaged the computer used to divert water from the Sacramento River to the local farms.	Insider attack. The person responsible for the computer systems.
4	* Mail/File server @ IT # Data Exfiltration	Data breach. Deleting and modifying information. An ex-employee was arrested on account of a computer crime.	Credential theft of another employee because of no 2FA, No password update policy, and no routine checks.

*	*Target @Domain # Attacker Action	Impact	Attack success Reason
5	* SCADA/HMI @ OT # Data Exfiltration	Data breach. The attack caused over \$30,000 in remediation costs. Whilst this attack had no consequences on the security and reliability of the Bowman Avenue Dam, it points to the vulnerabilities of critical water infrastructures, which are often monitored and controlled through unsafe web applications.	The control system was attached to the Internet via a cellular modem but was directly Internet accessible and not protected by a firewall or authentication access controls.
6	* Multiple @ IT and OT # Unauthorized Changes	Data manipulation. The attack disabled the communication between utilities and their data collection network, the organizations had to resume manual data gathering.	No implementation of access control policies and revoking access rights when someone is laid off.
7	* Multiple @ IT and OT # Unauthorized access	Control manipulation. The incident resulted in the exfiltration of 2.5 million unique records and the manipulation of chemicals and flow rates.	Internet-facing servers and applications, such as the payment management application here, should be connected to the SCADA.
8	* SCADA @ OT # Data Exfiltration	Data breach. No details of the key findings have been disclosed.	No information disclosed.
9	* Routers @ OT # Unauthorized access	Bandwidth theft. The hack was believed to be an opportunistic action to steal valuable internet bandwidth, resulting in the authority's cellular data bill soaring from an average of \$300 a month to \$45,000 in December 2016 and \$ 53,000 in January 2017.	The use of hard-coded credentials by the router manufacturer and the failure of the water authority to install the patches proved to be major contributors to this incident.

*	*Target @Domain # Attacker Action	Impact	Attack success Reason
10	* Account DB @ IT # Unauthorized access	Financial impact. The diverted refunds totaled over £500,000 and bitcoin was purchased.	No proper background checks were performed on the Employees of the partner CRM company.
11	* SCADA/HMI @ OT # Cryptojacking	Resource theft. The investigation classified nearly 40% of the traffic as related to mining operations, causing a 60% surge in the overall bandwidth consumption.	No Proper Early detection of the crypto-jacking malware.
12	* Info system @ IT # Ransomware	Data loss. The virus encrypted files and data, The authority had multiple layers of protection in place, including firewalls and antivirus/malware software. Yet, their IT system has proven to be penetrable.	No Proper Security team monitoring the Utility 24/7.
13	* Database @ IT and OT # Ransomware	Denial of access. Staff of the Utility were not able to access Technical data.	Not Disclosed.
14	*Database, SCADA @ IT and OT # Ransomware	Data loss. The attack compromised the computer systems controlling pumping stations and water quality testing and payment operations of the Utility. Approximately \$ 600,000, was paid to the attackers.	Old And Vulnerable Hardware. Lack of professionals needed to secure their IT and OT systems. No Basic Cyber-security Training.
15	* SCADA @ OT # Unknown	Not Disclosed	Not Disclosed

The major contribution towards documenting the cyber threat incidents targeting Water CPS was provided in the study [40]. The literature has systematically compiled and presented valuable insights into historical cyber-attacks on water systems. The paper conducted a thorough literature review, offering a comprehensive evaluation of the

existing state of cybersecurity for cyber-physical systems within the water sector. It categorized information related to multiple cyber-attack detection mechanisms, highlighting key aspects such as test beds, simulations, and datasets utilized in cybersecurity studies. Moreover, the literature review furnished an insightful overview of essential topics, including CPS, security measures for CPS protection, the SWaT, and the water distribution testbed (WADI). It also critically assessed the limitations of current cybersecurity solutions. This wealth of information serves as a valuable guide for this work. In summary, this literature review provides crucial insights into various dimensions of cyber-attacks, detection mechanisms, test beds, and existing cybersecurity limitations. In the other literature study [17], 15 incidents were reviewed, analyzed, and categorized according to the situation, response, remediation, and lessons learned, which are included in the table below with the other incidents that occurred. This study sheds light on the general architecture of SCADA, A Typical Architecture in Water Systems, and ISA-62443 Zoned Architecture, explaining the components in detail. These findings are instrumental in shaping the foundation of this work.

RQ1.3: What detection mechanisms are proposed for water systems against security threats?

The authors in the literature study [6] provided a comprehensive survey for the common cyber-physical attacks and common detection mechanisms for the water distribution system (WDS) in specific. The comparison of attacks and detection methods with a focus on ideas, methods, evaluation results, advantages, and various other factors. The study compared different cyber-physical attacks and detection algorithms and concluded the research by explaining that no optimal detection algorithm exists for identifying all the attacks.

The next table 5 provides information about the various detection methods proposed over the years, categorizing the type of the method proposed, results of the study, Plus points of using the detection method and disadvantages/limitations of the method. The Results are categorized or marked based on the detection of all the possible attacks on CPS mentioned in the cyber attack model constructed by authors in the study [34].

In the literature study [34], the authors tried to build attack models categorizing different types of cyber-physical attacks by identifying cyber-physical system components that respond to attacks. Further in the study [34], they propose nine different types of attacks on the components of CPS of Water Distribution System (WDS) such as PLC units, Sensors, and SCADA system. A simple and Trivial detection method was proposed. This Detection framework follows an object-oriented programming approach, wherein the implementation encompasses the representation of nine distinct attacks through individual classes. Within each class, specific attributes are defined, and a straightforward method is employed to ascertain various attack features, including the nature of the attack and its time duration. The major disadvantage of the detection method is the attack model

Table 5. Detection methods

Study	Type of Method	*Results, @Advantages	Restrictions
[34]	Simple Algorithm	*CPS is highly impacted during the attacks. @Testing of detection algorithms can be done using attacks proposed	Poor detection capabilities
[3]	Anomaly behaviour detection algorithm using combination of Statistical, ANN and PCA	* Identifies all the mentioned attacks without delay. @ Good performance	Can provide false Detection's
[13]	Anomaly behaviour detection algorithm using combination of Actuator rules, data verification and optimization	* Identifies all the mentioned attacks without delay. @ Good performance	The optimization algorithm is less effective compared to the Data Verification algorithm.
[28]	Anomaly behaviour detection algorithm using combination of three modules	* Identifies all the mentioned attacks without delay. @ Good performance	Takes lot of time for Detection of attacks
[4]	Anomaly behaviour detection algorithm using combination of four modules	* Identifies all the mentioned attacks without delay. @ Good performance	Lacks to recognize multiple compromised components at same time
[18]	Model-based Fault detection method with three phase approach	* Detects all the labeled simulated attacks . @ Good performance	Less reliable due to the impact of uncertain sensor data noise on result accuracy
[15]	Anomaly behaviour detection algorithm using LSTM-RNN and Cumulative Sum method	* Detects all the attacks and the attacked sensor . @ High accuracy	The method is tested and trained with a small sensor dataset of the SWaT Testbed.
[19]	Anomaly behavior detection algorithm using Deep Neural Network	* Detects all the 36 attacks on SWaT . @ High accuracy	Low performance and insensitive to subtle variations in both data and actuators.
[2]	Anomaly behaviour detection algorithm using One-class neural network	* Detects all the 36 attacks on SWaT . @ Good detection performance	Exhibits some detection delay in certain attack scenarios.
[8]	Simple Mathematical Method	* Detects all the attacks. @ Works as an Effective Method	There is a need for future extensions by developing fingerprinting for wireless networks.

was applied to simple CPS with one of each component, such as pump, tank, valve, and multiple but few actuators, which is not the case with real CPS, which is more complex.

The study [3] propose a detection algorithm to identify local anomalies affecting each sensor individually and global anomalies affecting multiple sensors simultaneously. The algorithms include 3 layers Simple statistical detection layer, an Artificial Neural network (ANN) layer, and a Principal Component Analysis (PCA) layer. Initially, the statistical approach identifies outliers within each sensor by comparing data values against established high and low boundaries for normal operations. However, this method may be susceptible to detecting false outliers. Subsequently, an Artificial Neural Network (ANN) model is employed, trained to recognize patterns in normal operations and forecast potential anomalies for each individual sensor. Despite its effectiveness, there is a challenge related to overfitting. Lastly, the Principle Component Analysis (PCA) technique addresses the high-dimensional nature of combined sensor data and facilitates

the detection of anomalies occurring simultaneously across multiple sensors[3].

In the study [13], the authors present an algorithm designed to detect malicious attacks by verifying sensor data integrity and actuator rule adherence, aiming to identify anomalies in the data. Following this, they employ an optimization approach to extract a low dimensionality of sensor data, effectively segregating it from all SCADA data measurements. This detection mechanism follows the combination of three algorithms working one after the other. At First Actuator Rules, the verification algorithm and data verification algorithms check the integrity of the SCADA measurements. At last, the Optimization Algorithm is run to detect the complex CPS attacks that were not detected by previous algorithms.

In the study [28], the algorithm to detect anomalies was presented, which included three modules: control rule and consistency module, pattern recognition module, and hydraulic and system relationships module. The control rule and consistency module checks the data consistency with specified control rules mentioned in the data set of control rules. The pattern recognition module contains different patterns for hydraulic parameters. The hydraulic parameters are developed based on datasets from cyber-attacks [28]. The other module Hydraulic and system relationships module, is constructed based on the relationship of WDS components; the calculated values of these components are then compared with collected data of WDS components to detect the attacks. As explained in the study [28], Each module within the three operates independently to identify attacks. The culmination of the attack detection results is derived by integrating the outcomes from all three modules, given they are connected through logical statements [28].

The authors of the study [4] propose a combined methodology of four modules to detect cyber-physical attacks. The first module verification of the actuator Rules module ensures that the operations of valves and pumps follow the right control rule based on the observed water levels in every tank. The next module monitors the actuator and sensors based on calculating upper and lower boundaries; if the value is more or less from the upper and lower boundary, the value will be considered an outlier [4]. In the subsequent module ANN, The dataset will be employed to train the Artificial Neural Network (ANN) model, enabling the prediction of future observations related to tank level data, pressure, and pumping flow rate. Finally, the PCA module involves remapping multi-dimensional sensor data onto new axes known as Principal Components (PCs). This separation into two datasets occurs based on variance, with one dataset having maximum variance classified as normal data, while the other dataset exhibiting the lowest variance is classified as an anomaly [4].

In the study [18], the authors propose the Model-based Fault detection method with a three-phase approach; in the first phase, the demand is estimated based on part of the SCADA readings; in the second phase, a hydraulic model is used for checking whether the SCADA systems hydraulic data corresponds to the estimated demand and in third

phase A multilevel classification approach is then implemented to classify the obtained errors into outlier and normal errors. This model uses a physically-based water hydraulics simulation model (EPANET) to detect cyber-attacks on water distribution systems. The algorithm introduced in this study [18] assesses normal errors against errors generated in the presence of attacks in SCADA readings to uncover potential cyber-attacks. While the results demonstrate the algorithm's capability to detect attacks from SCADA readings, its reliability is compromised due to the impact of sensor data noise on result accuracy. The results achieved by the proposed algorithm portray that it is capable of achieving the best-known performance when tested on the data published in the BATtle of the Attack Detection ALgorithms (BATADAL) competition (<http://www.batadal.net>) [18].

The authors in the study [15] propose an unsupervised learning approach using Long Short Term Memory Recurrent Neural Network (LSTM-RNN) to train the dataset of SWat Testbed and predict the attacks. A short explanation of RNN is, that it is a type of Deep Neural Network in which it feeds the output layer as input to the next layer. The LSTM algorithm addresses the vanishing gradient issue in RNN, enhancing the model's learning capabilities. LSTM incorporates a memory block comprising an input gate, a forget gate, and an output gate, thereby augmenting the model's capacity to learn data and make predictions. The Cumulative Sum method serves as a statistical technique for determining high and low boundaries. Initially, LSTM-RNN undergoes training on data and predicts the output. Subsequently, the Cumulative Sum method compares the predicted and actual sensor values. The method proposed in the study [15] could detect the anomalies and attacks in CPS and identify the sensor that was attacked with high accuracy. The small limitation of the study was that it was trained with a small sensor dataset of the SWaT Testbed [15].

The research [19] presented a novel intrusion detection approach employing deep neural networks based on machine learning. A notable strength of their methodology lies in its unsupervised operation, eliminating the need for labeled attack samples during training and relying solely on normal data. However, the detection performance exhibited limitations, reflected in an F score of 0.80281. This outcome is attributed to the method's insensitivity to subtle variations in both data and actuators, particularly in situations characterized by on-and-off dynamics. In the context of cybersecurity, the study underscores the significance of addressing the challenges associated with detecting nuanced variations in system behavior. The unsupervised nature of the approach holds promise for scenarios where labeled attack data is scarce, but there remains a need for further refinement to enhance sensitivity to subtle anomalies in both data and actuator states. This insight contributes to the ongoing discourse on improving intrusion detection systems, acknowledging the nuanced nature of cyber threats in cyber-physical systems[19].

The authors in the study[2] introduced an unsupervised one-class neural network, leveraging the advantage of minimal hyperparameters, simplifying the training and tuning processes. Despite its computational simplicity, the method attains commendable

detection performance. However, it is noteworthy that the overall efficacy does not reach the levels achieved by supervised deep learning methods. Furthermore, the method displays a tendency for detection delays in specific attack scenarios.

In the study [8], authors designed jamming attacks that block the communication channels to disrupt the communication between physical processes and PLCs. These jamming attacks are intended to exert control or inflict damage on a SWaT testbed. The outcomes of the literature work [8] reveal that the SWaT testbed exhibited a response and had adverse effects on water overflow. Moreover, the authors in the study [8] propose a straightforward detection method for identifying attacks by comparing measurement values with their properties. However, it is suggested that this detection method requires enhancement in future work to fortify the security of both the physical layer and the network layer.

The Study [8] used the detection mechanism proposed in the study [7], in which authors propose the method to detect cyber attacks on Water Treatment plants using Process Invariants. In their study [7], the authors proposed and tested a detection mechanism based on invariants derived from the physical design of the Cyber-Physical System (CPS). This mechanism aims to identify anomalies in the underlying process. Experiment results demonstrate the method's effectiveness in detecting attacks that, if undetected, could significantly impact the process behavior in an undesirable manner.

A "process invariant," or simply an invariant, refers to a mathematical relationship among the "physical" and "chemical" properties controlled by one or more Programmable Logic Controllers (PLCs) [8]. At any given time instant, a suitable set of these properties constitutes the observable state of the Secure Water Treatment (SWaT) system [8]. For instance, in a water treatment plant, such a relationship could involve the correlation between the water level in a tank and the flow rate of incoming and outgoing water for that tank. These properties are measured by sensors during CPS operation and captured by PLCs at predetermined time instants. The recorded measurements are often stored in a historian workstation for subsequent analysis[7].

In 2017, the BATADAL competition, held at a California conference, aimed to develop attack detection algorithms for identifying cyber-physical attacks in Water Distribution Systems (WDS). The study [35] evaluates various detection methods employed by seven participating teams in BATADAL, using metrics like detection time and the capability to identify compromised components.

The first team utilized a feature extraction method based on mean and covariance calculations, coupled with a rainforest algorithm for data classification into normal and abnormal categories. The second team employed district-metered areas to reduce data dimensionality and recurrent neural networks for classification and attack prediction. The third team proposed a method verifying rule operation integrity, classifying data through a deep neural algorithm known as a convolutional variational autoencoder [35]. The fourth team's approach involved checking the integrity of SCADA data and actuators' rules,

incorporating an optimization algorithm to minimize computation time. The fifth team designed a three-layer model for attack detection, incorporating outlier detection, artificial neural network classification, and anomaly determination based on principle component analysis [35]. The sixth team introduced a method with three modules to assess control rules, data integrity, and component relationships within the WDS. The seventh team presented a model based on EPENANT to simulate a WDS, comparing actual water system data with the simulation model to detect attacks. Results indicated that all teams successfully detected cyber-physical attacks, with the seventh team emerging as the overall winner. However, the algorithms were trained on medium-sized WDS, and the study [35] recommends considering large-sized WDS in future research.

In the study [10], authors proposed a novel Intrusion Detection System (IDS) for the SWaT Dataset, employing a hybrid DT. Enhanced anomaly detection was provided by leveraging inherent system dynamics and real-world data insights, and a granular attack localization feature was included, which allowed pinpointing security threats at the physical component level [10]. A simple thresholding mechanism was implemented to demonstrate DT's security capabilities and the IDS capability of DT was leveraged by comparing the water level output of the model with the physical water level using thresholding [10].

3.4 Identified gaps

Despite the comprehensive literature review conducted on cybersecurity threats targeting water systems and the proposed detection mechanisms, several significant gaps remain in the existing research. These gaps necessitate further investigation and exploration to advance the understanding and implementation of cybersecurity measures for water CPS. Several critical gaps have been identified that warrant further investigation:

- **Integration of Advanced Technologies:** Many existing studies focus on traditional detection methods and simplistic algorithms. There is a need to explore the integration of advanced technologies such as artificial intelligence (AI), machine learning (ML), and anomaly detection algorithms to enhance the accuracy and effectiveness of cybersecurity measures for water CPS.
- **Adaptability to Dynamic Threat Landscape:** Cyber threats targeting water systems are continually evolving, necessitating adaptive and resilient cybersecurity solutions. Future research should focus on developing dynamic detection mechanisms capable of adapting to the changing threat landscape and mitigating emerging cyber risks effectively.
- **Consideration of Human Factors:** Human error and insider threats pose significant challenges to the security of water CPS. Research efforts should explore the

role of human factors in cybersecurity and develop strategies to address human-centric vulnerabilities, such as training programs, awareness campaigns, and user behavior analysis.

- **Scalability and Resource Constraints:** Many proposed detection mechanisms may not be scalable or practical for deployment in real-world water infrastructure due to resource constraints and operational limitations. Future research should consider the scalability, cost-effectiveness, and feasibility of implementing cybersecurity solutions in diverse water system environments.
- **Interdisciplinary Collaboration:** Cybersecurity for water CPS requires interdisciplinary collaboration between cybersecurity experts, water engineers, policy-makers, and stakeholders. There is a need for greater synergy and collaboration between these domains to develop holistic and integrated cybersecurity strategies that address the unique challenges of water infrastructure protection.

3.4.1 Challenges in the Proposed Current System

- **Detection Accuracy:** One of the primary challenges in the proposed current system is achieving high detection accuracy. Many existing detection mechanisms rely on traditional methods or simplistic algorithms, which may not effectively detect sophisticated cyber-physical attacks targeting water systems.
- **Scalability:** Another challenge is the scalability of the proposed detection mechanisms. As water systems vary in size and complexity, it is essential to develop scalable solutions that can be tailored to different environments without compromising effectiveness or performance.
- **Real-time Response:** Ensuring real-time response capabilities is crucial for mitigating cyber threats effectively. However, existing systems may face challenges in providing timely responses to detected threats, particularly in large-scale water CPS deployments.

3.4.2 Why Digital Twin Technology?

DT technology presents a promising approach to addressing the identified challenges in securing water CPS against cyber threats. This Section elaborately discovers the various roles of DT in the context of cybersecurity while providing the answer to research question **RQ1: What role can Digital Twin play in improving the security posture of Water CPS?**. By creating virtual replicas of physical water systems, DTs offer several advantages:

- **Predictive Capabilities:** DTs enable predictive simulation, allowing for the identification of potential vulnerabilities and attack scenarios before they occur in the physical environment. This proactive approach enhances cybersecurity by preemptively addressing threats [16].
- **Real-time Monitoring:** Integrating IoT sensors into DTs enables real-time monitoring and analysis of water system data. This continuous monitoring facilitates early detection of anomalies and potential security breaches, enabling rapid response and mitigation efforts [32].
- **Simulation and Testing:** DTs provide a controlled environment for simulating and testing various security measures, such as intrusion detection systems and access controls. This allows for a thorough evaluation of security protocols before implementation in the physical water infrastructure [32].
- **Deception Decoy Implementation:** Leveraging DTs as deception decoys can divert cyber attackers' attention away from the actual water infrastructure, providing valuable insights into threat actors' tactics and intentions while safeguarding critical assets.

3.4.3 Why Blockchain?

The incidents observed in various locations and years underscore the critical need for robust cybersecurity measures to protect water infrastructure against cyber threats [40]. From unauthorized access to SCADA systems to sophisticated ransomware attacks, these incidents highlight the vulnerabilities present in water systems and the potential consequences of security breaches. One of the key challenges highlighted in these incidents is ensuring data integrity within the control systems of water infrastructure. Attackers often exploit vulnerabilities to manipulate data, compromise system integrity, and disrupt operations. Traditional security mechanisms may fall short in preventing unauthorized tampering with critical system properties, raising concerns about the reliability and trustworthiness of data used for decision-making in water CPS. This is where blockchain technology emerges as a crucial component in enhancing cybersecurity for water infrastructure. Blockchain, as a decentralized and immutable ledger, offers a transparent and tamper-proof record of transactions and data exchanges. By leveraging blockchain, water utilities can establish a secure and verifiable audit trail for all interactions within the CPS, including data transmission between components and the DT.

In the context of DT-based security mechanisms, blockchain provides several advantages:

- **Data Integrity Assurance:** Blockchain's inherent immutability ensures that pre-defined properties of components stored within the ledger remain unchanged and

tamper-proof. This prevents malicious actors from altering critical system parameters, thereby enhancing the integrity of data used by the DT for decision-making.

- **Secure Data Transmission:** Blockchain facilitates secure and transparent data transmission between components of the water CPS and the DT. Each transaction is cryptographically secured and verified by network participants, ensuring that only authorized and authenticated data is passed to the DT.
- **Immutable Audit Trail:** Blockchain maintains a chronological and immutable record of all transactions and data exchanges, enabling comprehensive auditing and traceability of system activities. This audit trail enhances accountability and transparency, allowing stakeholders to track the provenance of data and identify any unauthorized changes or anomalies.
- **Resilience to Cyber Attacks:** The decentralized nature of blockchain makes it resistant to single points of failure and tampering. Even in the event of a cyber attack targeting specific nodes or components, the integrity of data stored on the blockchain remains uncompromised, ensuring the continued functionality and reliability of the DT-based security mechanisms.

In summary, blockchain technology serves as a foundational layer for ensuring data integrity, transparency, and resilience in DT-based security solutions for water CPS. By integrating blockchain into the architecture of the DT, water utilities can establish a trusted and secure framework for defending against cyber threats and safeguarding critical infrastructure assets.

3.5 Summary

The systematic literature review conducted in this work aimed to explore the landscape of cybersecurity threats and detection mechanisms targeting water cyber-physical systems (CPS). The review encompassed a comprehensive analysis of existing research literature, addressing the following research questions:

RQ1.1: Security Threats: The review identified a wide range of security threats targeting water systems, including cyber-physical attacks aimed at disrupting services, compromising control equipment, and compromising communication networks. These threats pose significant risks to public health, environmental well-being, and the reliability of critical water infrastructure. Incidents such as the Maroochy Water Services incident in Australia (2000) and the Tehama-Colusa Canal incident in the U.S. (2007) underscored the potential risks posed by third-party contractors and insider threats, highlighting the need for robust security measures to mitigate these risks. **RQ1.2: Vulnerabilities Exploited:** Vulnerabilities exploited in water CPS include weaknesses in control systems, communication protocols, software vulnerabilities, and human-centric factors such

as insider threats and human error. Understanding these vulnerabilities is crucial for developing effective cybersecurity measures to protect water infrastructure from cyber threats. The incidents observed, such as the Bowman Avenue Dam incident in New York (2013) and the Onslow Water and Sewer Authority incident in North Carolina (2018), emphasized the critical importance of addressing vulnerabilities at both the system and personnel levels to prevent unauthorized access and data breaches. **RQ1.3: Detection Mechanisms:** Various detection mechanisms proposed for water systems included anomaly detection algorithms, intrusion detection systems, and model-based fault detection methods. The SLR highlighted studies that leveraged machine learning techniques, such as recurrent neural networks and deep neural networks, to identify cyber-physical attacks and anomalous behavior in water infrastructure. Additionally, the review identified the use of DTs as a promising approach for enhancing cybersecurity through predictive simulation, real-time monitoring, and deception decoy implementation.

The SLR paved the way to collect and compare (shown in the table 5 the best detection mechanisms proposed specifically for the SWaT architecture, which can be used as Anomaly Detection System in our proposed architecture of utilizing DT as an attack deception mechanism. The proposed studies [8] Anomaly behavior detection algorithm using LSTM-RNN and Cumulative Sum method, [2] Anomaly behavior detection algorithm using Deep Neural Network, [19] Anomaly behavior detection algorithm using One class neural network, [15] Simple mathematical method and [7] detection mechanism using process invariants were tested on the attacks mentioned in the SWaT dataset(explained in section 5.4) and showed the capability of detecting all the thirty-six attacks from the dataset. The study [10] proposed a novel Intrusion Detection System (IDS) for the SWaT Dataset, employing a hybrid DT. A simple thresholding mechanism was implemented to demonstrate DT's security and IDS capabilities.

The review highlights the importance of robust cybersecurity measures for safeguarding water CPS against evolving cyber threats. It underscores the potential of advanced technologies such as AI, ML, and DTs in enhancing the resilience and security of water infrastructure. However, several gaps and challenges remain, including the need for adaptive detection mechanisms, consideration of human factors, and interdisciplinary collaboration to address the complex cybersecurity landscape of water systems.

4 Use Case and Attack Scenarios

In the context of cybersecurity, a use case refers to a specific scenario or situation in which a system, process, or technology is applied to address a particular need or achieve a desired outcome. Use cases help to illustrate how a solution or approach can be practically implemented to solve real-world problems or fulfill specific requirements. In the context of water infrastructure and cybersecurity, use cases may involve the application of DT technology, IRPs, or other security measures to enhance the resilience and security of

water systems against cyber threats.

An attack scenario, on the other hand, describes a potential cyber threat or security breach that could occur within a system or network. It outlines the methods, tactics, and objectives of potential adversaries and the vulnerabilities and weaknesses in the system that could be exploited. Organizations can better understand their security risks by analyzing attack scenarios and developing effective strategies to mitigate and respond to cyber threats. In the context of water infrastructure cybersecurity, attack scenarios may involve various threat actors targeting critical components of water systems, such as sensors, actuators, control systems, and communication networks, with the goal of disrupting operations or causing harm.

4.1 Use Case 1: Water Supply Infrastructure and Cybersecurity

In an increasingly digitized world, the intersection of water infrastructure and cybersecurity has emerged as a critical area of concern. As societies rely heavily on water systems for essential services and daily life, ensuring the security and resilience of these systems against cyber threats is paramount.

Overview of water systems

Water systems encompass various infrastructure components designed to manage, distribute, and treat water for domestic, industrial, and agricultural purposes. These systems range from simple groundwater wells and distribution networks to complex water treatment plants and reservoirs. Understanding the intricacies of these systems is essential for comprehending the challenges posed by cybersecurity threats.

Different Types of Water Systems

Within the domain of water infrastructure, various types of systems serve distinct functions and cater to diverse user needs. These include but are not limited to:

Potable Water Systems: Responsible for providing safe drinking water to communities, potable water systems comprise treatment facilities, distribution networks, and storage reservoirs.

Wastewater Systems: Tasked with collecting, treating, and disposing of wastewater, these systems prevent environmental contamination and ensure public health.

Industrial Water Systems: Supporting manufacturing processes and industrial operations, industrial water systems require specialized treatment and management to meet specific quality and quantity requirements.

Agricultural Irrigation Systems: Facilitating crop irrigation and agricultural activities, these systems utilize water resources efficiently while addressing sustainability concerns.

Critical Infrastructure Components

Central to the water infrastructure discussion, cybersecurity is the critical component un-

derpinning water systems' functioning. These components, including sensors, actuators, control systems, and communication networks, play vital roles in monitoring, controlling, and maintaining water infrastructure operations. However, their interconnected nature and reliance on digital technologies make them susceptible to cyber threats, necessitating robust cybersecurity measures to safeguard against potential disruptions and breaches.

The suggested approach strengthens defense mechanisms against cyberattacks by integrating blockchain technology with DT technology to improve the cybersecurity of water supply infrastructure. The solution aims to increase the resilience and security of water infrastructure by building virtual replicas of crucial water system components and integrating them with blockchain for secure communication and data integrity. The need for proactive cybersecurity measures is made clear by thoroughly understanding various water system types and vulnerabilities. The suggested solution can proactively detect, mitigate, and respond to cyber threats by identifying potential attack vectors and threat actors targeting water infrastructure, protecting the integrity and dependability of water supply services. Furthermore, the use case emphasizes how crucial it is to protect critical infrastructure elements against malicious intrusions by highlighting the components—such as sensors, actuators, control systems, and communication networks vulnerable to cyber threats. Stakeholders can monitor and control these components in real-time, identify anomalies, and implement incident response strategies to mitigate potential disruptions and breaches by implementing DTs integrated with blockchain.

This use case clarifies the complex interactions between cybersecurity and water infrastructure overall. It emphasizes the importance of finding creative ways to safeguard vital water resources and guarantee that society will always have access to them.

4.2 Use Case 2: Digital Twins in Water Supply

As the water sector embraces digitization, adopting DT technology emerges as a promising approach to enhance efficiency, resilience, and sustainability. DTs, virtual replicas of physical assets, systems, or processes, have gained traction in the water sector for their ability to simulate, monitor, and optimize water infrastructure operations. DTs enable stakeholders to gain insights, make informed decisions, and predict outcomes in a virtual environment by creating digital counterparts of real-world assets, such as treatment plants, distribution networks, and reservoirs. This section delves into DTs in the water sector, exploring their role, benefits, and challenges in transforming water infrastructure management.

Role of Digital Twins in Water Supply

DTs play multifaceted roles across various water infrastructure lifecycle management stages in the water sector. DTs facilitate data-driven decision-making, asset performance optimization, and predictive maintenance, from design and construction to operation

and maintenance. DTs empower water utilities and stakeholders to proactively address challenges, improve operational efficiency, and enhance service delivery by providing real-time insights into system behavior.

In the context of enhancing the cybersecurity of water supply infrastructure, the integration of DTs with blockchain technology presents a compelling solution to fortify defense mechanisms against cyber threats. By creating virtual replicas of critical water system components and leveraging blockchain for secure communication and data integrity, the proposed solution aims to enhance the resilience and security of water infrastructure operations.

The proposed solution involves integrating digital twin technology with blockchain as an attack deception mechanism. By creating virtual replicas of critical water system components and incorporating blockchain for secure communication and data integrity, the solution aims to deceive potential attackers and protect telemetry data from insider threats. DTs play a pivotal role in this context by enabling real-time monitoring, anomaly detection, and predictive analytics, empowering stakeholders to identify and mitigate potential cyber threats proactively. Moreover, DTs facilitate the development and implementation of incident response strategies tailored to the unique challenges of water infrastructure cybersecurity. By simulating various attack scenarios and evaluating their potential impact on water system operations, DTs enable stakeholders to refine their incident response protocols and enhance their readiness to address emerging cyber threats effectively. Leveraging the insights provided by DTs, water utilities and stakeholders can optimize their cybersecurity posture and minimize the risk of disruptive cyber incidents, thereby safeguarding the integrity and reliability of water supply services.

Overall, integrating DTs with blockchain technology represents a synergistic approach to enhancing the cybersecurity and resilience of water supply infrastructure. By harnessing the power of DTs to simulate, monitor, and optimize water system operations, stakeholders can proactively address cybersecurity challenges and ensure the uninterrupted delivery of safe and reliable water services to communities.

4.3 Attacker Model and Scenarios

An attacker model defines the various scenarios and capabilities of potential adversaries in a cybersecurity context. This work adopted a systematic approach to assessing the system's security, leveraging an established attack model [25] tailored specifically for CPSs. This attack model enables the generation of attack procedures and functions aimed at a particular CPS, in our case, the SWaT testbed. In the case of the SWaT Testbed [25], three distinct attacker settings were considered:

- **Attacker A:** This attacker possesses access to the local plant communication network.

- **Attacker B:** While not physically on-site, this attacker is in close proximity to the plant.
- **Attacker C:** This attacker is physically present on site and has direct access to the devices.

Attacker Goals: Regardless of the specific setting, the primary objective of the attacker remains consistent: to manipulate the plant's normal operations. For instance, an illustrative attack aimed to overflow the raw water tank within the SWaT system. While such an attack may not result in significant damage, it serves as a representative example of the type of manipulation requiring full control over sensors and actuators in an Industrial Control System (ICS) environment [25].

Methods of Attack:

The attackers, designated as A, B, and C, employ various tactics to fully manipulate communication within the L0 ring or L1 networks of the SWaT system. These tactics include:

- **Man-in-the-middle Attacks:** Attackers can insert themselves between two parties within the network, such as two Programmable Logic Controllers (PLCs), enabling them to eavesdrop on all exchanged sensor and command data [25].
- **On-the-fly Data Modification:** Utilizing specialized tools like Ettercap⁴, attackers can rewrite sensor or command values in real-time, facilitating the alteration of system behavior [25].

The CPS attack model represented as a sextuple (M; G; D; P; S₀; S_e), encompasses various components: M, a set of attack procedures; G, a subset of attacker intents; D, the domain model derived from the CPS; P, a finite set of attack points; and S₀ and S_e, infinite sets representing the possible start and end states relevant to the attacker [14]. Attack points within the CPS could include physical elements or entry points within the communication network linking sensors, actuators, Programmable Logic Controllers (PLCs), and the SCADA system [14].

Our analysis underscores the vastness of the potential attack landscape, stemming from the flexibility to modify attack methods (M), attack points (P), and CPS states (S₀ and S_e) [14]. The sheer scale of the attack space highlights the intricate nature of securing CPS environments and emphasizes the importance of adopting proactive security measures, such as implementing IRPs leveraging innovative approaches like DTs for attack deception.

⁴<https://www.ettercap-project.org/>

4.3.1 Attacks Targeting P1 stage

Table 6 presents a comprehensive overview of attacks directed at P1 stage, spanning various types, including Single Stage Single Point Attacks (1, 2, 3, 34, 36), Single Stage Multi Point Attacks (21, 35), Multi-Stage Single Point Attacks (26), and Multi-Stage Multi Point Attacks (30). The "No" column represents the attack number which is noted from the Dataset explained in 5.4. Given that the DT representation is confined to the P1 stage of the SWaT system, the assessment of the proposed DT deception mechanism primarily concentrates on the impact of attacks targeting stage one.

Table 6. Attacks targeting P1 Stage process of SWaT

Attack No	Start time	End Time	Attack Point	Start state	Attack	Actual change	Expected Impact
1	10:29:14 AM	10:44:53 AM	MV-101	MV-101 is closed	Open MV-101	Yes	Tank overflow
2	10:51:08 AM	10:58:30 AM	P-102	P-101 is on where as P-102 is off	Turn on P-102	Yes	Pipe bursts
3	11:22:00 AM	11:28:22 AM	LIT-101	Water level between L and H	Increase by 1 mm every second	No	Tank Underflow; Damage P-101
21	6:30:00 PM	6:42:00 PM	MV-101, LIT-101	MV-101 is open; LIT-101 between L and H	Keep MV-101 on continuously; Value of LIT-101 set as 700 mm	Yes	Tank overflow
26	5:04:56 PM	5:29:00 PM	P-101, LIT-301	P-101 is off; P-102 is on; LIT-301 is between L and H	P-101 is turned on continuously; Set value of LIT-301 as 801 mm	Yes	Tank 101 underflow; Tank 301 overflow
30	3:47:40 PM	4:07:10 PM	LIT-101, P-101, MV-201	P-101 is off; MV-101 is off; MV-201 is off; LIT-101 is between L and H; LIT-301 is between L and H	Turn P-101 on continuously; Turn MV-101 on continuously; Set value of LIT-101 as 700 mm; P-102 started itself because LIT301 level became low	Yes	Tank 101 underflow; Tank 301 overflow
33	2:21:12 PM	2:28:35 PM	LIT-101	Water level between L and H	Set LIT-101 to above H	No	Tank underflow; Damage P-101
34	5:12:40 PM	5:14:20 PM	P-101	P-101 is on	Turn P-101 off	Yes	Stops outflow
35	5:18:56 PM	5:26:56 PM	P-101; P-102	P-101 is on; P-102 is off	Turn P-101 off; Keep P-102 off	Yes	Stops outflow
36	10:16:01 PM	10:25:00 PM	LIT-101	Water level between L and H	Set LIT-101 to less than LL	No	Tank overflow

4.3.2 Attack scenarios

In this section, we build an example scenario from Table 6 to explain how the proposed solution DT integrated with Blockchain as an attack deception mechanism would mitigate the attack in its early stages, reducing the further damage to the actual physical process. We also discuss how the role-based Incident Response Playbook might be effective in this scenario.

Attack Scenario 1:

Attack number: 1;

Attack point: MV101;

Start state: MV101 is closed;

Attack: Open MV101;

Expected Impact and Actual change: Tank Overflow, Yes;

In this attack scenario, the attacker tried to manipulate the component from the P1 stage process of SWaT MV101 to change its property state from Closed to Open state, which resulted in the overflow of tank T101, impacting the whole physical process. Considering the scenario, if the attacker can manipulate the physical process in Level O of the network architecture of SWaT, it is safe to assume that the attacker has already compromised the SCADA, HMI, and Workstations placed in Level 3 (Detailed explanation on High-level network architecture of SWaT is provided in Section 2.2. As mentioned in the proposed solution 8, DT is placed strategically between Level 1 and Level 2 of the architecture, acting as an attack deception mechanism with open ports; it is safe to assume two actions from the attacker. First, The attacker might assume DT is a real physical process as DT is a perfect mimic with fake telemetry data, which allows real-time simulation and tries to manipulate it. In this case, the DT was able to mitigate the attack by acting as an attack deception mechanism. Second, The attacker might ignore the open ports and try to manipulate the real physical component; in this case, the anomaly detection system, which is placed in level 2 of the architecture, can detect the behavioral anomaly and raise an alert, and CPS operator can follow the IRP proposed and take necessary actions mentioned in IRP. Once the initial actions are taken, an attacker will lose access to the real physical component and be led to DT automatically as the network connections from Level 1 to Level 3 are completely shut off. As DT is placed between these levels, an attacker would manipulate the DT components, assuming they are real physical components. In this case, DT and IRP were able to mitigate the attack at an early stage, preventing further damage and collecting the much-needed attacker information while safeguarding the real physical components.

Attack Scenario 2: Insider Attack ⁵

As discovered in the literature review (Section 3), various attacks in the past occurred via insider threats; we built the insider attack scenario to showcase the potential of the proposed Blockchain-integrated solution in defending the DT data in such case. Consider an example of an insider attack scenario where the CPS operator knows the architecture of CPS, where DT acts as an attack deception mechanism and tries to manipulate and

⁵<https://www.geeksforgeeks.org/what-is-insider-attack/>

change the fake telemetry data supplied to the DT to damage the DT model components. In such scenarios, the Ethereum smart contracts serve as a safeguarding layer. Any tampering with the property values triggers a record within the blockchain, providing an immutable and auditable log of the attempted manipulation. Through the utilization of blockchain technology, the integrity and authenticity of the telemetry data passed to the DT are preserved, thus bolstering its resilience against insider threats.

4.4 Summary

In this section, we explored the intersection of cybersecurity with water infrastructure, emphasizing the importance of use cases and attack scenarios in addressing potential threats. Use cases illustrated practical applications of DT technology and IRPs to enhance the security and resilience of water systems against cyber threats. By creating virtual replicas of critical water system components and integrating them with blockchain for secure communication and data integrity, stakeholders can proactively detect, mitigate, and respond to cyber threats. Attack scenarios highlighted various threat scenarios, including insider attacks and manipulations of critical components, underscoring the need for robust defense mechanisms. Leveraging DTs as attack deception mechanisms and blockchain for data protection, stakeholders can mitigate attacks at early stages, safeguarding the integrity and reliability of water supply services. Overall, the use case and attack scenarios shed light on the intricate interplay between cybersecurity and water infrastructure, emphasizing the importance of innovative solutions to safeguard critical water resources in an increasingly digitized world.

5 Solution Design

As discovered in the section 3, there have been multiple attack detection algorithms and mechanisms proposed in the past, and no study introduced us to the attack deception mechanism and standard procedures that can be followed in an attack scenario and even remediate the attack in the preliminary stages of the cyber attack. In this work, we propose the attack deception mechanism leveraging DT Technology integrated with Blockchain and Incident response strategies utilizing the IRP. The SWaT Testbed (explained in Section 2.2, is considered the base architecture for water CPS. The SWat testbed attack dataset, explained in section 5.4 was considered, as the cyberattacks are performed in the Water system's physical process layer, which coincides with this work of implementing the DT integrated with Blockchain as an attack deception mechanism assuming the attacker is already in the water systems network and can modify properties of the system remotely. In further subsections, the architecture of utilizing the DT as an attack deception mechanism in SWaT architecture is explained in detail.

5.1 Proposed HLC Portion Architecture Of SWaT With Digital Twin

The High-Level Control Portion (HLC) architecture of SWaT is explained in Section 2.2, and detailed architecture is depicted in figure 3, where multiple stages of the process are depicted. In this work, we propose the deception technique considering the single stage P1 of the SWaT Testbed Process, and as shown in the figure 8, the deception technique by creating the DT of the Single process stage is implemented. As explained in the 2.2, the PLC and the physical process are implemented in Level 0 and Level 1 of the architecture and SWaT, and further, the SW1 switch is connected to the SCADA via SW2 in Level 2 of control portion architecture. As shown in figure 8, in the connection between SW1 and SW2 i.e, Level 1 and Level 2, an Anomaly Detection System (ADS) is being proposed for implementation, which is further connected via C1 to the virtual DT environment which is a perfect mimic of the physical process. This digital twin can be a deception mechanism if the anomaly detection system detects any anomalies in the physical process. The DT is connected to the physical process via a broken channel connection to fake portray the real-time data exchange between the physical process and the DT replica. But in reality, the DT is fed with fake telemetry data, which will be created in a way that is very similar to the original physical data.

5.2 Proposed Digital Twin Design

This section explains in detail the proposed DT design. In this work, the DT is built based on the exact physical design and specifications of the P1 stage process of the SWaT Testbed. Figure 9 shows the simplified view of the P1 stage process of SWaT. Overall, Stage P1 controls the water inflow to be treated by opening or closing a valve that connects the inlet pipe to the raw water tank. Water from the raw water tank is pumped to the P2 stage process.

In Detail, the raw water is let into Tank T101 by turning the motorized valve MV101 to an "ON/OFF" state. The motorized valve acts as an actuator. The flow sensor FIT101 monitors MV101 and gives the reading of the water inflow rate in L/s. The tank system comprises a singular compartment denoted as T101, possessing a volumetric capacity of 1800 cubic meters, a height measuring 1.36 meters, and a diameter of 1.38 meters. The tank T101 is monitored by a level sensor called LIT101, which gives the T101 level in m3. Further, Pump P101 acts as an actuator with a predefined capacity measured in m3, feeding the water from T101 to the next process stage, P2, which is being monitored by the FIT201 flow sensor. The DT in the scope of this work is constructed with fake telemetry data, which will be very similar to the original data from the physical device. This is done to protect the original configuration as the DT is proposed as the attack deception mechanism to keep the attacker engaged in the DT.

The operational dynamics of the tank DT serve to maintain the integrity of sensory data, safeguarding against unauthorized alterations. However, the system's security ex-

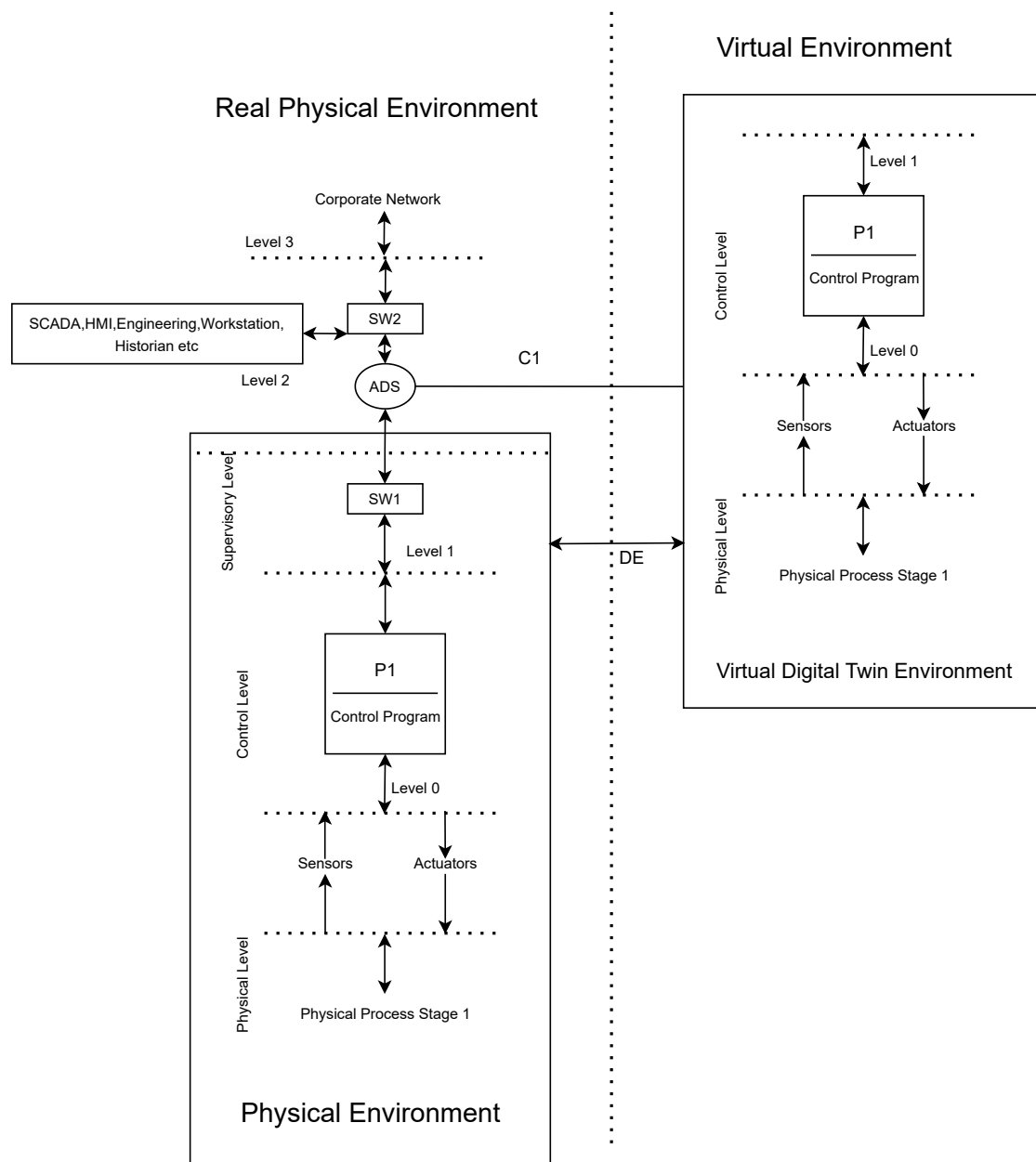


Figure 8. Proposed High-Level Control Portion Architecture of SwaT with DT as Deception Technique. C1 denotes communication with DT in Level 2. ADS denotes Anomaly Detection system and DE denotes broken connection for fake Data exchange between physical Twin and Digital Twin

tends beyond preventing unauthorized modifications to sensory data; it also encompasses the protection of control logic, which may be vulnerable to exploitation through attacks

on the programmable logic controller. To effectively simulate and address such attack scenarios, the DT must encompass not only the sensory data but also the control logic of the system [10].

Consider a hypothetical scenario wherein the system has a predefined high-level setting of 800m for the water level. In this scenario, a valve, MV101, is compromised, resulting in an influx of water into the system, even when the water level exceeds the high limit. Consequently, the compromised MV101 fails to respond appropriately, leading to overflowing the tank [10]. In such a scenario, the DT must accurately replicate the overflow that would occur in the physical system. To ensure that the system operates within the desired parameters, the DT must incorporate appropriate checks and control mechanisms akin to those present in the physical system[10].

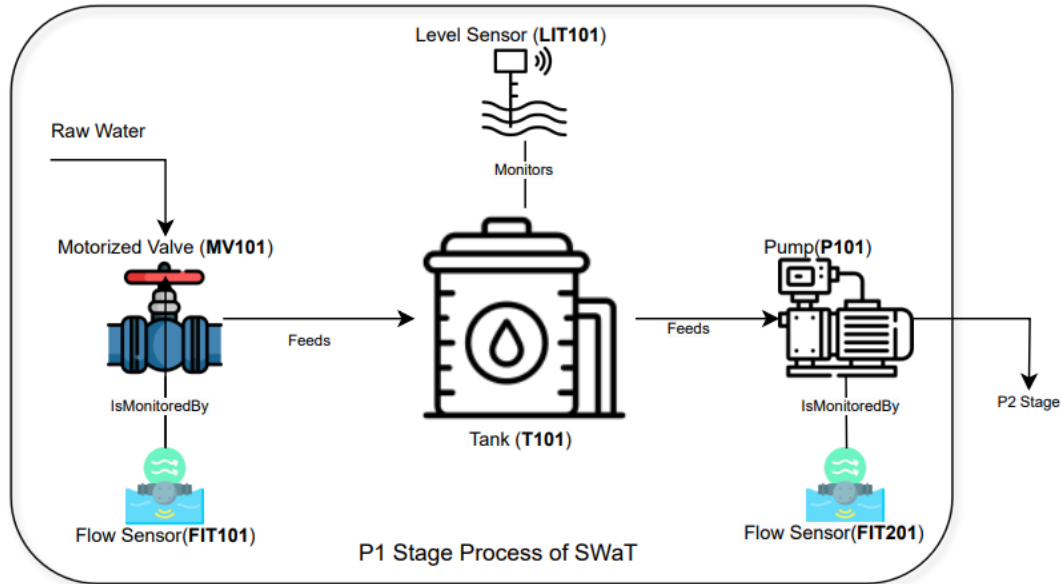


Figure 9. P1 Stage Process of SWaT

5.3 Ethereum Based Solidity Smart Contract

In the context of this work, Ethereum smart contracts are utilized to fortify the defense mechanisms of the DT model against insider attacks. These smart contracts are developed using Solidity and deployed on the Ethereum Sepolia test network. Within the scope of the solution design, the smart contracts are employed to store and manage the telemetry data property values associated with the Tank T101 and level sensor LIT101, which are integral components of the DT model.

Consideration is given to potential insider attack scenarios, as evidenced by historical incidents (refer to tables 3 and 4), wherein malicious actors attempt to manipulate the property values of the DT. In response, security personnel, responsible for safeguarding the Water CPS, establish predefined threshold values for key parameters, such as the water level in Tank T101. For instance, a threshold range of 2m³ to 5m³ may be set to prevent underflow or overflow situations.

In the event of an unauthorized attempt by a CPS operator or employee to modify the property values within the DT, acting as an attack deception mechanism, the Ethereum smart contracts serve as a safeguarding layer. Any tampering with the property values triggers a record within the blockchain, providing an immutable and auditable log of the attempted manipulation. Through the utilization of blockchain technology, the integrity and authenticity of the telemetry data passed to the DT are preserved, thus bolstering its resilience against insider threats.

5.4 SWaT Dataset

The SWaT dataset [14] comprises a comprehensive time-series record spanning a duration of 11 days and encompasses a total of 51 attributes. These attributes include 26 continuous values representing sensor readings alongside 25 discrete-state actuators, with states such as "on," "off," and transitional states. Crucially, the dataset encompasses essential information pertaining to various process variables, such as water flow rates, pH levels, temperature, pressure, and valve positions, all meticulously sampled at one-second intervals. The dataset is distinctly categorized into two main splits: one containing solely normal operational data, while the other incorporates 36 instances of attack implementations. In the scope of this work, as shown in table 6, only ten attack instances targeting the P1 stage process of SWaT were considered. In detail, the dataset comprises 890,298 samples representative of normal operations and 54,621 samples indicative of attacks, thereby revealing an inherently imbalanced distribution.

As explained in the subsection 2.2 and portrayed in figure 2, the water treatment system encompasses six sequential stages, each serving a specific function in the purification process. These stages are designated as follows: P1, responsible for the intake and storage of raw water; P2, tasked with the addition of chemicals to enhance water quality; P3, dedicated to ultra-filtration processes; P4, facilitating dechlorination through the utilization of ultraviolet lamps; P5, focused on reverse osmosis filtration; and finally, P6, serving as the storage and distribution point for the treated water.

The primary rationale for incorporating the SWaT testbed attack dataset into this investigation stems from the alignment between the perpetrated cyberattacks, and this work's focus. Specifically, the attacks targeted the physical process layer of water systems, which directly corresponds to the objective of implementing DT technology as an attack deception mechanism. It is assumed that the attacker has already gained access to the water systems network and can remotely manipulate system properties, thus

underscoring the relevance of the dataset to the research objectives at hand.

5.5 Anomaly Detection System

In the scope of this work, an Anomaly detection system is utilized and placed in Level 2 of the proposed architecture as shown in fig 8, which was proposed in various studies in the past. The proposed studies [8] Anomaly behavior detection algorithm using LSTM-RNN and Cumulative Sum method, [2] Anomaly behavior detection algorithm using Deep Neural Network, [19] Anomaly behavior detection algorithm using One class neural network, [15] Simple mathematical method and [7] detection mechanism using process invariants were tested on the attacks mentioned in the SWaT dataset(explained in section 5.4) and showed the capability of detecting all the thirty-six attacks from the dataset. The study [10] proposed a novel Intrusion Detection System (IDS) for the SWaT Dataset, employing a hybrid DT. A simple thresholding mechanism was implemented to demonstrate DT's security and IDS capabilities. The proposed detection mechanisms mentioned have the potential to detect anomalies and raise alerts when utilized in Level 2 of the architecture, as shown in the implementation architecture 10. The role and utilization of the ADS proposed in the various studies in past are discussed in Section 4 and Section 7.1 considering an attack scenario.

5.6 Proposed Incident Response Playbook Design

An incident response playbook defines common processes or step-by-step procedures needed for your organization's incident response efforts in an easy-to-use format [37]. This work introduces a series of tailored IRP designed to address the specific cyberattacks observed on the SWaT Testbed, as outlined in Table 6. These playbooks are developed primarily on simplicity and accessibility, ensuring they are easily understandable by water system operator personnel and other stakeholders. Emphasizing generic applicability, the playbooks are crafted to be compatible with various types of water systems, thereby enhancing their utility and effectiveness across different operational contexts. There are various types of playbooks and various ways to implement the playbook.

In the scope of this work, the IRP will be utilized and implemented in combination with the Azure DT model, which acts as a cyber-attack deception mechanism for the evaluation of the proposed solution. The playbook design is created by following the Cybersecurity and Infrastructure Security Agency (CISA) ⁶ guidelines of creating the IRP. The proposed playbook is created based on the phases defined by the National Institute of Standards and Technology (NIST)⁷ lifecycle framework. The NIST incident

⁶https://www.cisa.gov/sites/default/files/2024-03/Federal_Government_Cybersecurity_Incident_and_Vulnerability_Response_Playbooks_508C.pdf

⁷<https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-61r2.pdf>

response lifecycle comprises four primary phases: Preparation, Detection and Analysis, Containment, Eradication and Recovery, and Post-Event Activity.

- **Phase 1 - Preparation :** This initial phase encompasses the organizational efforts to equip itself for incident response, including establishing necessary tools, resources, and team training. Additionally, preventive measures are undertaken to mitigate the occurrence of incidents.
- **Phase 2 - Detection and Analysis:** Identifying and evaluating incidents accurately pose significant challenges for organizations, as outlined by NIST. This phase involves the meticulous detection and assessment of incidents to ascertain their nature and scope.
- **Phase 3 - Containment, Eradication, and Recovery :** Focusing on minimizing the impact of incidents and mitigating service disruptions, this phase is dedicated to containing the incident and implementing measures for eradication and subsequent recovery.
- **Phase 4 - Post-Event Activity :** An often overlooked but crucial phase, the post-event activity, involves the thorough analysis of the incident and the effectiveness of the response efforts. The objectives include reducing the likelihood of future incidents and enhancing the efficacy of future incident response endeavors.

As the IRP is designed for the Water CPS, Role-based Containment and Eradication actions become necessary. Thus, the IRP is constructed and implemented in two parts accordingly, considering One part for the role of **CPS Operator** with basic knowledge of the process and very little knowledge of security and the other part for the role of **internal or external cyber-security personnel** with the required knowledge of cyber-security and process. The IRP is created assuming that the CPS is being monitored by any Cybersecurity product, such as a Security Information and Event Management (SIEM)⁸ Tool, and the CPS operator is provided with a certain device to receive alerts from ADS. The IRP is specifically designed to defend Water CPS from attacks targeting the P1 stage process mentioned in the table 6 collected from the SWaT Dataset 5.4.

5.7 Summary

The Solution Design section presents a comprehensive strategy for bolstering the cybersecurity of Water CPS through innovative approaches. Central to this strategy is integrating DT as an attack deception mechanism within the existing architecture of the SWaT Testbed as explained in Section 6.1 and Section 5.1. By meticulously replicating

⁸<https://www.ibm.com/topics/siem>

the operational dynamics of critical CPS components, such as the P1 stage process, the DT serves as a strategic decoy, diverting potential attackers and safeguarding against cyber threats as shown in the Section 5.2. Furthermore, Ethereum-based Solidity smart contracts are employed to fortify the integrity of telemetry data associated with the DT, ensuring resilience against insider attacks. Overall, this section exhibits the designs produced in this work and answers two main research questions, **RQ2: What architectural frameworks facilitate the integration of DTs into Water CPS security?** in Section 6.1, Section 5.1 and **RQ3: How can Blockchain technology enhance the security of DTs in Water CPS?** in Section 5.3. In addition to DTs, the solution design encompasses deploying anomaly detection systems proposed in past studies to identify and mitigate cyber threats in real-time. Tailored incident response playbook designs are crafted to provide a systematic framework for incident detection, containment, eradication, and recovery, empowering water system operators and cybersecurity personnel to respond effectively to cyber threats. The implementation of the proposed solution in this work is provided in Section 6.

6 Implementation

In this section, we delve into the practical aspects of implementing the proposed solution, which was discussed in the section 5. The implementation process involves creating and integrating DTs into the Azure Digital Twin (ADT) framework, tailored specifically for the components of the P1 stage process of the SWaT testbed, Ethereum-based solidity smart contract deployment, and the creation of the incident response playbook. By incorporating DTs into the incident response playbook, we aim to bolster the resilience of the water treatment system against cyber threats while facilitating efficient monitoring and control.

6.1 Proposed Attack Deception Architecture Of SWaT with Digital Twin

This section sheds light on the prototypical implementation of the proposed conceptual framework of HLC portion architecture (explained in section 6.1) by generating a perfect Virtual mimic i.e., DT environment of the physical P1 stage process of SWaT. The Flow sensors, Motorized valve, Pump and Level Sensors can be modeled and analyzed through the twinned environment. The implementation leverages the Microsoft Azure Digital Twins service (ADT)⁹, which orchestrates the modeling of DTs within a cloud-based framework, adhering to a Platform-as-a-Service (PaaS) architecture. This service

⁹<https://azure.microsoft.com/en-us/products/digital-twins>

facilitates the monitoring of physical twins within a simulated environment, as depicted in Figure 10, outlining the architectural framework employed for the simulation.

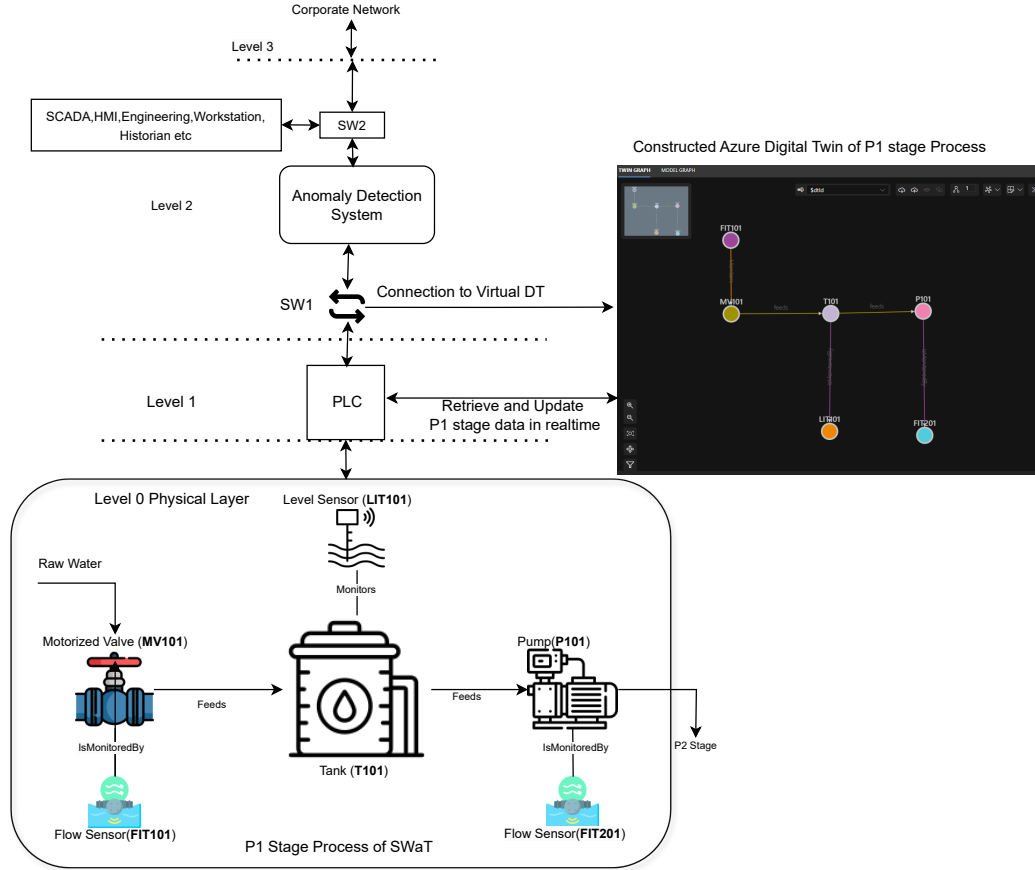


Figure 10. Implementation of Proposed Attack Deception Architecture

As shown in figure 10, the implementation consists of four major parts:

- **Physical P1 Stage process:** As explained in the section 5.2, consists of all the physical components and the operations of the P1 stage process, which is placed in Level 0 of the SWaT network architecture.
- **Constructed Digital Twin:** The DT, which mimics the physical process, is created in Azure DT Cloud framework and is placed strategically in Level 1 of the network architecture and is connected via Switch SW1 to the Anomaly Detection to act as cyber-attack deception in the attack scenario. The DT is also connected to the physical components' Program logic control (PLC) to retrieve and update the DT data and operations in real-time.

- **Anomaly detection system:** The anomaly detection system is placed in Level 2 of the network architecture, which is connected to SCADA and HMI in Level 3 via the SW2 switch. The anomaly detection system is explained in detail in section 5.5.
- **SCADA, HMI, Engineering, workstation, Historian etc :** The SCADA and components are placed on the border of Level 2, which are connected via the SW2 switch with the corporate network in Level 3 and also with the ADS. A brief explanation of the SCADA and its components is described in section 2.2.

Further, the proposed implementation architecture can be explained using an example attack scenario. Consider an attack scenario where the attacker is inside the Water Supply System network and tries to modify the property or value of the physical component to obstruct normal operational behavior. The DT is placed strategically in close proximity and connected to the physical system via a different network, and thus, the change in the operational behavior or property is mimicked in the DT in real-time. This change in the behavior would be flagged as an anomaly by ADS, and switch SW1 will direct all the connections from SCADA to the DT Environment, which is a perfect replica of the real physical twin. Meanwhile, the attacker is changing the properties in the DT without knowing it is a virtual environment, making the DT act as an ideal attack deception mechanism, giving the edge to the Security team to collect IOCs for the future, reduce the damage to the real physical Water Supply System, and contain the attack. Closing the connections from level 3 to the physical system in level 0 has to be done manually by the CPS operator. The predefined step-by-step procedure of actions to be taken in this kind of scenario is further explained via the utilization of the Incident Response Playbook (implemented in section 7.1 and design explanation in section 5.6), which can be universally applied across all the Water CPS's.

6.2 Digital Twin Models

In the Azure DT framework, users have the ability to define the vocabulary necessary for constructing DTs that mirror the physical layer components. This capability is facilitated by creating models for each twin, capturing the essence of the corresponding physical element. Notably, ADT models are articulated in the Definition Language (DTD¹⁰), which operates on the JSON-LD standard. This structured approach enables users to delineate DT's intricate characteristics and behaviors, facilitating a comprehensive representation of the underlying physical infrastructure within the ADT ecosystem. For the P1 Stage process scenario of the SWAT, 6 Models were created to represent the involved components: 1. Flow Sensor1, 2. Flow Sensor2, 3. Level Sensor, 4. Motorized Valve, 5. Pump, 6. Tank.

¹⁰<https://learn.microsoft.com/en-us/azure/digital-twins/concepts-models>

6.2.1 Flow Sensor: FIT101

To represent a Flow sensor1 in the twin model, as Listing 1 shows, 1 property is defined as FlowRate in L/s. Here, the flow rate is simulated as a measurement of the flow rate by the sensor. As in the P1 stage process of the SWAT testbed, the flow sensor twin model is named FIT101 to mimic the physical twin. The relationship is defined as "Monitors" and the target as a "Motorized Valve," which explains the simulation that the flow sensor monitors the motorized valve and gives the readings in liters per second as defined in the property.

```
{
  "@id": "dtmi:com:example:FlowSensor1;1",
  "@type": "Interface",
  "displayName": "FIT101",
  "@context": "dtmi:dtdl:context;2",
  "contents": [
    {
      "@type": "Property",
      "name": "FlowRateLs",
      "schema": "integer"
    },
    {
      "@type": "Relationship",
      "name": "Monitors",
      "displayName": "Monitors",
      "target": "dtmi:com:example:MotorValve;1"
    }
  ]
}
```

Listing 1. Flow Sensor1 model defined in JSON-LD using DTDL specifications

6.2.2 Flow Sensor: FIT201

To represent a Flow sensor2 in the twin model, as Listing 2 shows, 1 property is defined as FlowRate in L/s. Here, the flow rate is simulated as a measurement of the flow rate by the sensor. As in the P1 stage process of the SWAT testbed, the flow sensor twin model is named FIT201 to mimic the physical twin.

```
{
  "@id": "dtmi:com:example:FlowSensor2;1",
  "@type": "Interface",
  "displayName": "FIT201",
  "@context": "dtmi:dtdl:context;2",
```

```

    "contents": [
      {
        "@type": "Property",
        "name": "FlowRateLs",
        "schema": "integer"
      }
    ]
  }
}

```

Listing 2. Flow Sensor2 model defined in JSON-LD using DTDL specifications

6.2.3 Level Sensor: LIT101

As Listing 3 shows, 1 property is defined as Level in m to represent a Level Sensor in the twin model. Here, the level of the tank is simulated by the sensor as a measurement of the water level in a tank. As in the P1 stage process of the SWAT testbed, the flow sensor twin model is named LIT101 to mimic the physical twin.

```

{
  "@id": "dtmi:com:example:LevelSensor;1",
  "@type": "Interface",
  "displayName": "LIT101",
  "contents": [
    {
      "@type": "Property",
      "name": "levelm",
      "schema": "double"
    }
  ],
  "@context": "dtmi:dtdl:context;2"
}

```

Listing 3. Level Sensor model defined in JSON-LD using DTDL specifications

6.2.4 Motorized Valve: MV101

To represent a Motorized Valve in the twin model, as Listing 4 shows, one property, namely Functional State, has been defined, which explains if the motorized valve is "ON" or "OFF" state. As in the P1 stage process of the SWAT testbed, the motorized valve twin model is named MV101 to mimic the physical twin. The relationship is defined as

"Feeds" and the target as "Tank", which explains the simulation that a motorized valve feeds the water to the tank.

```
{
  "@id": "dtmi:com:example:MotorValve;1",
  "@type": "Interface",
  "displayName": "MV101",
  "contents": [
    {
      "@type": "Property",
      "name": "state",
      "schema": "string"
    },
    {
      "@type": "Relationship",
      "name": "feeds",
      "displayName": "Feeds",
      "target": "dtmi:com:example:Tank;1"
    }
  ],
  "@context": "dtmi:dtdl:context;2"
}
```

Listing 4. Motorized Valve model defined in JSON-LD using DTDL specifications

6.2.5 Pump: P101

To represent a Pump in the twin model, as Listing 5 shows, two properties have been defined: Duty, measured in m3/h, and Functional State, which explains if the pump is "ON" or "OFF" state. As in the P1 stage process of the SWAT testbed, the pump twin model is named P101 to mimic the physical twin. The relationship is defined as "IsMonitoredBy" and the target as "Flow Sensor2", which explains the simulation that the pump is being monitored by flow sensor 2.

```
{
  "@id": "dtmi:com:example:Pump;1",
  "@type": "Interface",
  "displayName": "P101",
  "contents": [
    {
      "@type": "Property",
      "name": "capacitym3",
      "schema": "double"
    },
    {

```

```

    "@type": "Property",
    "name": "state",
    "schema": "string"
  },
  {
    "@type": "Relationship",
    "name": "IsMonitoredBy",
    "displayName": "IsMonitoredBy",
    "target": "dtmi:com:example:FlowSensor2;1"
  }
],
"@context": "dtmi:dtdl:context;2"
}

```

Listing 5. Pump model defined in JSON-LD using DTDL specifications

6.2.6 Tank: T101

To represent a Tank in the twin model, as Listing 6 shows, one property has been defined, namely the tank's capacity, measured in m3. As in the P1 stage process of the SWAT testbed, the Tank twin model is named T101 to mimic the physical twin. Two relationships have been defined for the model. One the relationship is defined as "IsMonitoredBy" and the target as "Level Sensor", which explains the simulation that the level sensor is monitoring the Tank, and the other relationship, "Feeds," with the target as "Pump" explains the simulation that the Tank feeds the water to the Pump.

```

{
  "@id": "dtmi:com:example:Tank;1",
  "@type": "Interface",
  "displayName": "T101",
  "@context": "dtmi:dtdl:context;2",
  "contents": [
    {
      "@type": "Property",
      "name": "Capacitym3",
      "schema": "double"
    },
    {
      "@type": "Relationship",
      "name": "IsMonitoredBy",
      "displayName": "IsMonitoredBy",
      "target": "dtmi:com:example:LevelSensor;1"
    },
    {
      "@type": "Relationship",
      "name": "feeds",

```



```

        "displayName": "Feeds",
        "target": "dtmi:com:example:Pump;1"
    }

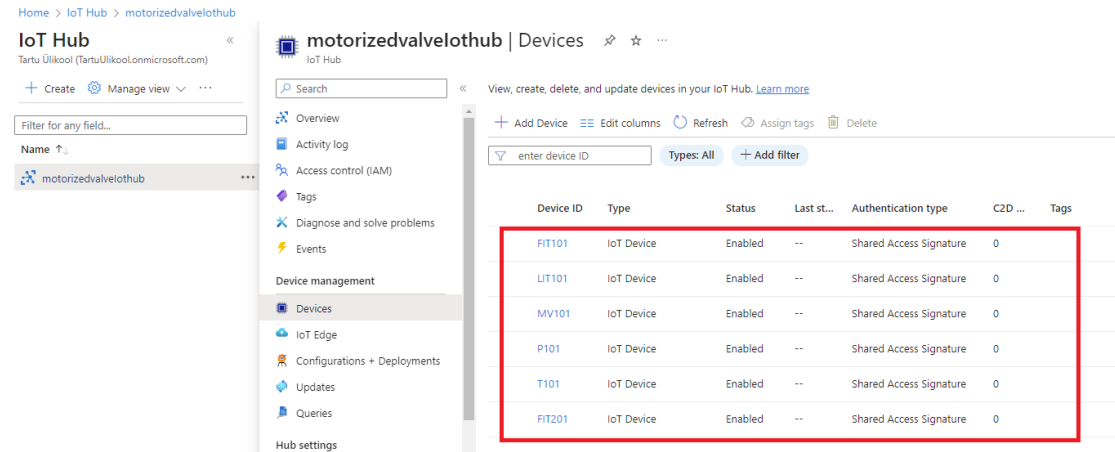
]
}

```

Listing 6. Tank model defined in JSON-LD using DTDL specifications

6.3 Digital Twin Simulators

The simulators are used to create mimic real-time data from the sensors and push it to the DT component. To simulate the real entities (pump, flow sensors, motorized valve, tank, and level sensor), *C#* console applications have been used. Six devices *FIT101*, *LIT101*, *MV101*, *P101*, *T101*, *FIT201* have been created under an instance to represent the components of the P1 stage process of SWaT Testbed architecture. The naming is done based on the real names of the components of SWaT Testbed architecture. As seen in the figure 11, the devices are subscribers to the data events pushed by the simulators to the Azure IoT Hub instance. These devices simulate real devices and interact with the device instances in the Azure portal cloud as shown in figure 11. The Level Sensor LIT101 simulator application is described below. The codebase for all other Data simulators is publicly available on GitHub¹¹.



Device ID	Type	Status	Last st...	Authentication type	C2D ...	Tags
FIT101	IoT Device	Enabled	--	Shared Access Signature	0	
LIT101	IoT Device	Enabled	--	Shared Access Signature	0	
MV101	IoT Device	Enabled	--	Shared Access Signature	0	
P101	IoT Device	Enabled	--	Shared Access Signature	0	
T101	IoT Device	Enabled	--	Shared Access Signature	0	
FIT201	IoT Device	Enabled	--	Shared Access Signature	0	

Figure 11. Instances Of Device Components in Azure IOT Hub

The simulator application replicates authentic sensory data in JSON format, which is subsequently transmitted to the corresponding Azure IoT hub devices for further

¹¹<https://github.com/Ojusvt/thesis-DigitalTwinsAttackDeception/tree/main/Simulator>

processing and utilization in updating the DTs. Within the Level Sensor simulation process context, the primary function employed is the invocation of the SimulateDeviceToSendD2cAndReceiveD2c function.

```
namespace LevelSensorSimulator
{
    public class Program
    {
        public static async Task Main(string[] args)
        {
            // Sample 1: Create device if you didn't have one in
            // Azure IoT Hub, FIRST YOU NEED SPECIFY connectionString
            // first in AzureIoTHub.cs
            //await CreateDeviceIdentity();

            // Sample 2: comment above line and uncomment following
            // line, FIRST YOU NEED SPECIFY connectingString and
            // deviceConnectionString in AzureIoTHub.cs
            await SimulateDeviceToSendD2cAndReceiveD2c();
        }

        public static async Task CreateDeviceIdentity()
        {
            string deviceName = "LIT101";
            await AzureIoTHub.CreateDeviceIdentityAsync(deviceName);
            Console.WriteLine($"Device with name '{deviceName}' was
                created/retrieved successfully");
        }

        private static async Task
            SimulateDeviceToSendD2cAndReceiveD2c()
        {
            var tokenSource = new CancellationTokencSource();

            Console.CancelKeyPress += (s, e) =>
            {
                e.Cancel = true;
                tokenSource.Cancel();
                Console.WriteLine("Exiting...");
            };
            Console.WriteLine("Press CTRL+C to exit");

            await Task.WhenAll(
                AzureIoTHub.SendDeviceToCloudMessageAsync(tokenSource
                    .Token),
                AzureIoTHub.ReceiveMessagesFromDeviceAsync(
                    tokenSource.Token));

            tokenSource.Dispose();
        }
    }
}
```

```

    }
}
}

```

Listing 7. Main Function to communicate with Level Sensor created in IoT Hub

The function depicted in Listing 7 does not require any parameters but invokes two asynchronous functions from the `AzureIoTHub` class. The first function, `SendDeviceToCloudMessageAsync`, illustrated in Listing 9, is responsible for transmitting telemetry data to the IoT Hub. Conversely, the second function, `ReceiveMessagesFromDeviceAsync`, depicted in Listing 10, handles the reception of any corresponding replies from the IoT Hub.

```

private static string iotHubConnectionString = @"HostName={hubName}
.azure-devices.net;SharedAccessKeyName=iothubowner;
SharedAccessKey={hubSharedAccessKey}";
private static string deviceConnectionString = $"HostName={hubName}.
.azure-devices.net;DeviceId={deviceName};SharedAccessKey={
deviceSharedAccessKey}";

```

Listing 8. connection strings to connect to azure cloud

The **`SendDeviceToCloudMessageAsync`** function executes its task through three sequential steps:

- **Instantiation of IoT Hub:** Initially, it instantiates a virtual device within the IoT Hub by utilizing a connection string termed `deviceConnectionString`, as illustrated in Listing 8. Within this string, `hubName` denotes the name assigned to the IoT Hub instance established in the Azure Portal. Furthermore, `deviceName` represents the unique identifier designated during the device's creation within the IoT hub (e.g., LIT101), while `sharedDeviceAccessKey` signifies the automatically assigned private key utilized to grant external system access to the device.
- **Creation of mock telemetry data:** Subsequently, a fabricated telemetry data object is generated for simulation purposes. A reading representing the water level (of double data type) is created in this instance. The message data is then transformed into a JSON object and dispatched to the designated device within the IoT Hub.
- **Await completion and iterative process:** Finally, the function awaits the completion of the asynchronous request and iterates the process at regular intervals, typically every minute. The duration of the delay interval can be adjusted as necessary using the `Task.Delay()` function.

```

public static async Task SendDeviceToCloudMessageAsync(
    CancellationToken cancellationToken)
{
    var deviceClient = DeviceClient.
        CreateFromConnectionString(deviceConnectionString);
    string id = deviceId;
    double levelm = 3;
    while (!cancellationToken.IsCancellationRequested)
    {
        var telemetryDataPoint = new
        {
            levelm = levelm,
        };
        var messageString = JsonSerializer.Serialize(
            telemetryDataPoint);
        var message = new Microsoft.Azure.Devices.Client.
            Message(Encoding.UTF8.GetBytes(messageString))
        {
            ContentType = "application/json",
            ContentEncoding = "utf-8"
        };
        await deviceClient.SendEventAsync(message);
        Console.WriteLine($"{DateTime.Now} > Sending message:
            {messageString}");

        //Keep this value above 1000 to keep a safe buffer
        //above the ADT service limits
        //See https://aka.ms/adt-limits for more info
        await Task.Delay(30000);
    }
}

```

Listing 9. Function Sending telemetry to virtual Level Sensor in IoT Hub

```

public static async Task ReceiveMessagesFromDeviceAsync(
    CancellationToken cancellationToken)
{
    try
    {
        string eventHubConnectionString = await
            IotHubConnection.GetEventHubsConnectionStringAsync(
                iotHubConnectionString);
        await using var consumerClient = new
            EventHubConsumerClient(
                EventHubConsumerClient.DefaultConsumerGroupName,
                eventHubConnectionString);
    }
}

```

```

        await foreach (PartitionEvent partitionEvent in
            consumerClient.ReadEventsAsync(cancelToken))
        {
            if (partitionEvent.Data == null) continue;

            string data = Encoding.UTF8.GetString(
                partitionEvent.Data.Body.ToArray());
            Console.WriteLine($"Message received. Partition:
                {partitionEvent.Partition.PartitionId} Data:
                '{data}'");
        }
    }
    catch (TaskCanceledException) { } // do nothing
    catch (Exception ex)
    {
        Console.WriteLine($"Error reading event: {ex}");
    }
}

```

Listing 10. Receiving response from IoT Hub

```

C:\Users\ojusvire\Downloads\ > Press CTRL+C to exit
4/19/2024 1:52:28 PM > Sending message: {"levelm":3}
Message received. Partition: 0 Data: '{"levelm":3}'
4/19/2024 1:52:58 PM > Sending message: {"levelm":3}
Message received. Partition: 0 Data: '{"levelm":3}'
4/19/2024 1:53:28 PM > Sending message: {"levelm":3}
Message received. Partition: 0 Data: '{"levelm":3}'
4/19/2024 1:53:58 PM > Sending message: {"levelm":3}
Message received. Partition: 0 Data: '{"levelm":3}'

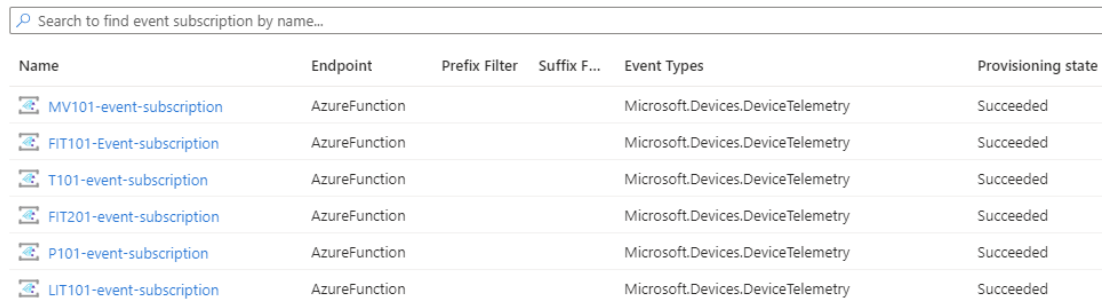
```

Figure 12. Level sensor simulator sending telemetry data

However, the second function `ReceiveMessagesFromDeviceAsync` creates a consumer client to consume any message (JSON Object) coming from the virtual LevelSensor device in IoT Hub. This function waits for any incoming messages from IoT Hub, converts them to strings, and logs them to the console window. This goes on as long as the process is not exited. Figure 12 shows typical message logs for the application.

6.4 Digital Twin Data Ingestors

The data ingestors utilized in the proposed solution are Azure function applications, which are responsible for retrieving data sent to the IoT Hub in the form of events generated by Data simulator applications. These Azure function applications undertake the analysis of the received data and subsequently update the DTs corresponding to components within the sensory data communication of the P1 stage in the proposed architectural framework. Functioning as event subscribers, these applications possess Azure Data Owner roles assigned to the IoT hub within the resource group established as the simulation layer on the Azure cloud platform. The Azure function applications designated for data ingestors within the proposed architecture are openly accessible on GitHub¹². Implementation for the Level Sensor-LIT101 ingestor application is illustrated here.









Name	Endpoint	Prefix Filter	Suffix F...	Event Types	Provisioning state
 MV101-event-subscription	AzureFunction			Microsoft.Devices.DeviceTelemetry	Succeeded
 FIT101-Event-subscription	AzureFunction			Microsoft.Devices.DeviceTelemetry	Succeeded
 T101-event-subscription	AzureFunction			Microsoft.Devices.DeviceTelemetry	Succeeded
 FIT201-event-subscription	AzureFunction			Microsoft.Devices.DeviceTelemetry	Succeeded
 P101-event-subscription	AzureFunction			Microsoft.Devices.DeviceTelemetry	Succeeded
 LIT101-event-subscription	AzureFunction			Microsoft.Devices.DeviceTelemetry	Succeeded

Figure 13. Event subscriptions in Azure IOT Hub for the devices

For our proposed DT, every event directed to the IoT Hub is accompanied by specific subscriptions, as depicted in Figure 13. Each subscription is linked to an Azure function responsible for handling the incoming device data and subsequently updating their corresponding DT representations. For example, the (LIT101-event-subscription) is for our Level Sensor LIT101 device on IoT Hub. The highlighted Azure function IoTHubToLIT101ADTFunction in Figure 14 is the ingestor function updating the Level Sensor LIT101 DT based on the telemetry data update events from the Level Sensor simulator. For developing the ingestor Azure functions for the proposed architecture, references have been taken from Azure's sample Github¹³ repository.

To illustrate the operation of the Level Sensor LIT101 Ingestor Azure function application, we initially execute the Level Sensor simulator, generating simulated telemetry data based on the technical specifications of the SWaT Testbed. This data is then transmitted to the Azure IoT Hub as events. Subsequently, the ingestor application is deployed directly within the Azure cloud under our designated resource group, with a subscription

¹²<https://github.com/Ojusvt/thesis-DigitalTwinsAttackDeception.git>

¹³<https://github.com/Azure-Samples/digital-twins-samples/>

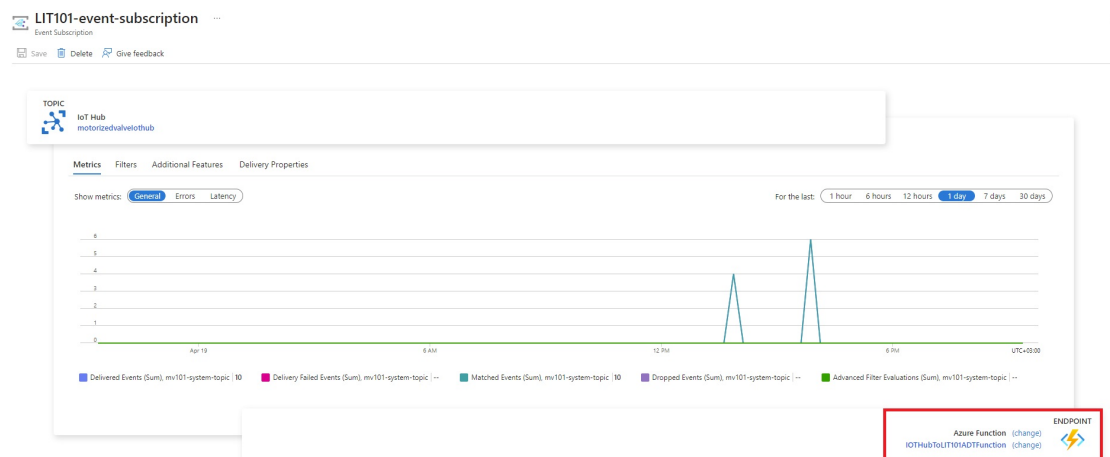


Figure 14. Azure function for the Level Sensor LIT101 in Azure IOT Hub

assigned to the Level sensor LIT101 device (LIT101-event-subscription). The Azure function IoTHubToLIT101ADTFunction intercepts these event triggers and processes the contained data. Upon processing, the DT on the ADT Explorer is updated accordingly. The successful updating of the DT is logged within the metrics for the ingestor Azure function IoTHubToLIT101ADTFunction. Figure 15 illustrates the data ingestion flow diagram.

Furthermore, the Level Sensor ingestor Azure function application is encapsulated within a C# class named IoTHubToLIT101ADTFunction, which encompasses a singular asynchronous task function. This function is responsible for orchestrating the logic associated with capturing event telemetry data originating from the pump simulator and effectuating updates to its DT counterpart within the ADT Explorer. The function encompasses four distinct tasks, as depicted in Listing 11.

- **Instantiation of ADT client:** Initially, the function instantiates an ADT client, establishing a connection with the virtual device event subscription previously configured within Azure. This process leverages a configuration variable, denoted as *"ADT_SERVICE_URL"*, which is defined within the Azure environment post-application deployment. This variable essentially encapsulates the URL of the Azure instance pertinent to our designated resource group.
- **Retrieval of telemetry data:** Subsequently, the function scrutinizes the EventGrid Data for telemetry information. Upon detection of such data, the telemetry is captured by deserializing it into a JSON object. In the case of the level sensor ingestor, pertinent telemetry metric such as the water level are extracted from the simulator.
- **Level Sensor data from smart contract:** Then it communicates with the level sensor smart contract to get the water level readings and compare them with the

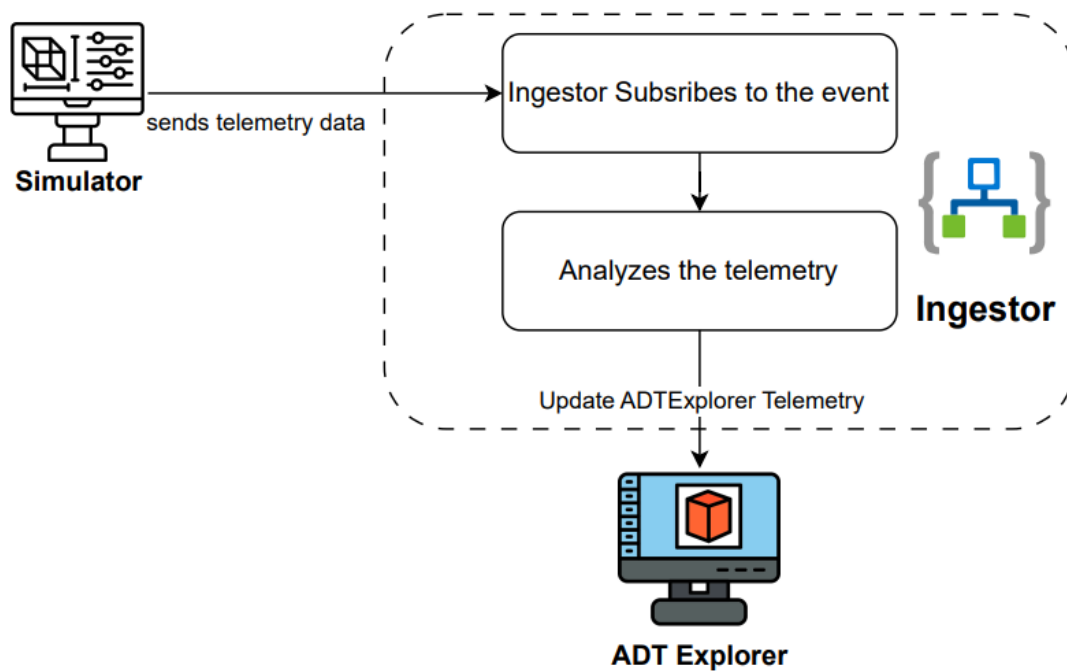


Figure 15. Data Ingestion by Azure Function Flow Diagram

level sensor simulator data. In case level sensor data falls inside the range, it proceeds to the next step of updating the data in the ADT; in another case, it throws a warning log message alerting the cybersecurity personnel that the data is being tried to manipulate.

- **Update of ADT Explorer twin:** Finally, armed with the freshly acquired values pertaining to the properties of the Level Sensor LIT101, the function undertakes the task of updating the corresponding LIT101 ADT data housed within the ADT Explorer. The updated property value and DT can be verified in the ADT explorer as shown in figure 16 and 17.

```

namespace LevelSensorIngestFunction
{
    [FunctionOutput]
    public class GetWaterLevelOutputDTO : IFunctionOutputDTO
    {
        [Parameter("int", "minLevel", 1)]
        public virtual BigInteger MinLevel { get; set; }
        [Parameter("int", "maxLevel", 2)]
        public virtual BigInteger MaxLevel { get; set; }
    }
}
  
```



```

public static class Function1
{
    // ADT Instance
    private static readonly string adtInstanceUrl = Environment.
        GetEnvironmentVariable("ADT_SERVICE_URL");
    private static readonly HttpClient httpClient = new
        HttpClient();

    // Contract addresses
    private static string selfAddress = "0
        xB7A5bd0345EF1Cc5E66bf61BdeC17D2461fBd968";
    private static string tankAddress = "0
        xa16E02E87b7454126E5E10d957A927A7F5B5d2be";
    private static string sepoliaApiKey = "373
        dcc9def4e4a97a73caec95874ca8c";

    [FunctionName("IOTHubToLIT101ADTFunction")]
    public static async Task Run([EventGridTrigger]
        EventGridEvent eventGridEvent, ILogger log)
    {
        log.LogInformation(eventGridEvent.Data.ToString());

        if (adtInstanceUrl == null) log.LogError("Application
            setting \"ADT_SERVICE_URL\" not set");

        try
        {
            //Managed Identity Credentials
            var cred = new DefaultAzureCredential();

            //Instantiate ADT Client
            var adtClient = new DigitalTwinsClient(new Uri(
                adtInstanceUrl), cred);

            // Log successful connection creation
            log.LogInformation($"LIT101 ADT client connection
                created!");

            //if we receive data
            if (eventGridEvent != null && eventGridEvent.Data !=
                null)
            {
                //Log the data
                log.LogInformation(eventGridEvent.Data.ToString()
                    );

                //Covert to json
                JObject LIT101Message = (JObject)JsonConvert.
                    DeserializeObject(eventGridEvent.Data.ToString

```

```

    ());

    //Get device data from object
    string LIT101Id = (String)LIT101Message["
        systemProperties"]["iothub-connection-device-
        id"];

    double levelm = (double)LIT101Message["body"]["
        levelm"];

    //Log the telemetry
    log.LogInformation($"Device: {LIT101Id} levelm: {
        levelm}");

    //Smart Contract code
    var web3 = new Web3($"https://sepolia.infura.io/
        v3/{sepoliaApiKey}");
    var tankABI = @"[{ ""inputs"": [], ""
        stateMutability"": ""nonpayable"", ""type"": ""
        constructor"", {"inputs"": [], ""name"": ""
        getWaterLevelThresholds"", ""outputs"": [{ ""
        internalType"": ""uint256"", ""name"": """", ""
        type"": ""uint256""}, {"internalType"": ""
        uint256"", ""name"": """", ""type"": ""uint256""
        }], ""stateMutability"": ""view"", ""type"": ""
        function""}]";

    // Initialize Tank contract
    var contract = web3.Eth.GetContract(tankABI,
        tankAddress);
    var waterLevel = contract.GetFunction("
        getWaterLevelThresholds");

    //Deserialize
    var tempLevels = await waterLevel.
        CallDeserializingToObjectAsync<
        GetWaterLevelOutputDTO>();

    //Log the values
    log.LogInformation($"Min water level: {tempLevels
        .MinLevel}");
    log.LogInformation($"Max water level: {tempLevels
        .MaxLevel}");

    int maxWaterLevel = (int)tempLevels.MinLevel;
    int minWaterLevel = (int)tempLevels.MaxLevel;

    if (levelm < minWaterLevel)

```

```

        {
            log.LogWarning($"Water level below threshold
                           detected!");
        }
        if (levelm > maxWaterLevel)
        {
            log.LogWarning($"Water level above threshold
                           detected!");
        }

        log.LogInformation($"Inforamtion from smart
                           contract analyzed!");

        // Update the digital twin explorer
        var updateLIT101TwinData = new JsonPatchDocument
        ();
        updateLIT101TwinData.AppendReplace("/levelm",
            levelm);

        await adtClient.UpdateDigitalTwinAsync(LIT101Id,
            updateLIT101TwinData).ConfigureAwait(false);
    }
}
catch (Exception ex)
{
    log.LogError($"Error in LIT101Id Ingest Function: {ex
        .Message}");
}
}
}
}
}

```

Listing 11. Azure function class for Level Sensor-LIT101 ingestor

6.5 Solidity Smart Contracts

To facilitate the analysis and governance of Tank T101 and Level Sensor LIT101 property values, Ethereum smart contracts have been devised utilizing Solidity and subsequently deployed on the Ethereum Sepolia test network¹⁴. The development, compilation, and deployment of these contracts have been conducted employing Hardhat¹⁵.

¹⁴<https://sepolia.dev/>

¹⁵<https://hardhat.org/docs>

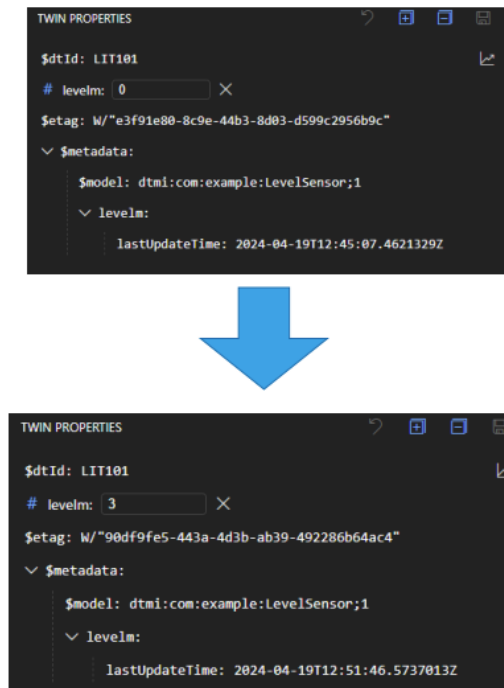


Figure 16. Verification of LIT101 Property update on ADT Explorer

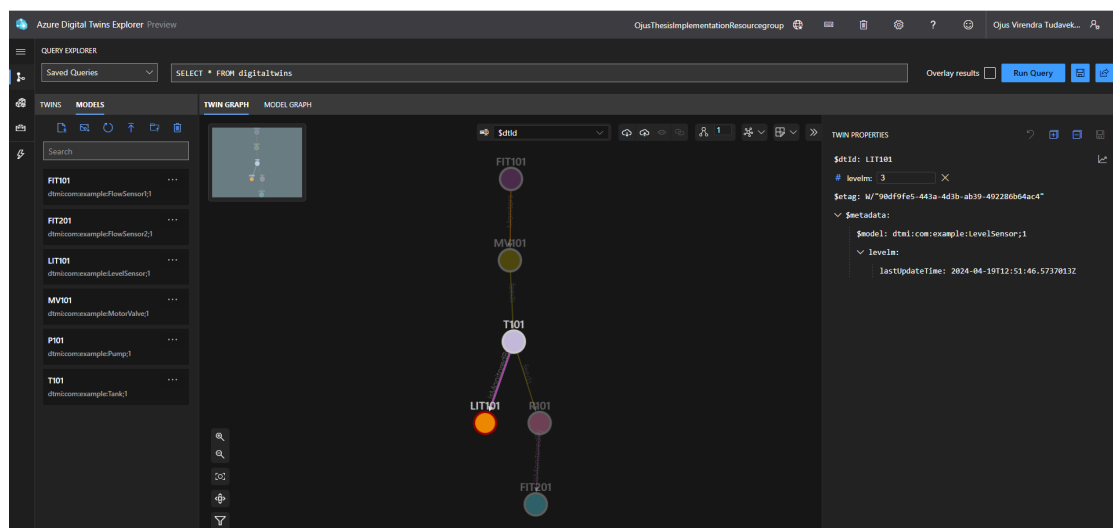


Figure 17. Verifying LevelSensor LIT101 DT update on ADTExplorer

6.5.1 Tank T101 Contract

As Listing 12 shows, there are two functions: (a) *setLevel* and (b) *getLevel* to interact with in the contract for a Tank. The system might trigger the WaterLevelSet event by calling the setLevel function which takes an integer parameter (waterLevel) to be set after successful analysis of the Water Level from the analyzer in the cloud. To retrieve the current water Level of the Tank T101.

```
//SPDX-License-Identifier: UNLICENSED
pragma solidity ^0.8.9;
contract Tank {

    int private waterLevel;

    event WaterLevelSet(address setter, int waterLevel);

    function setLevel(int _lvl) public {
        waterLevel = _lvl;
        emit WaterLevelSet(msg.sender, waterLevel);
    }

    function getLevel() public view returns (int) {
        return waterLevel;
    }

}
```

Listing 12. Smart Contract used for Tank T101 Digital Twin

6.5.2 Level Sensor LIT101 Contract

The level Sensor LIT101 returns the maximum and minimum water level thresholds in the current scenario in the cloud. These maximum and minimum threshold water level values are set to defend the P1 stage process from Insider attacks by validating the changes made unknown to the security admin and flagging the state for water level as malicious or not, which can help in defending water tank T101 from underflowing or overflowing of the tank. Listing 13 illustrates the Level Sensor LIT101 contract.

```
//SPDX-License-Identifier: UNLICENSED
pragma solidity ^0.8.9;

contract LevelSensor{
    uint256 minLevel;
    uint256 maxLevel;
```

```

    constructor () {
        minLevel =1;
        maxLevel =5;
    }

    function getWaterLevelThresholds() public view returns (uint256,
        uint256) {

        return (minLevel, maxLevel);
    }
}

```

Listing 13. Smart Contract used for Level Sensor LIT101 Digital Twin

6.5.3 Process Stage Contract

The Process Stage Contract assumes the responsibility of deploying additional contracts. It has been structured to serve as a parent component, incorporating T101 and LIT101 as its child components, to delineate the dependence of the level sensor on Tank T101 for precise water level measurement. Moreover, the Process Stage Contract governs the deployment of the LIT101 and T101 contracts. The method *getChildContractAddresses* is employed to retrieve the addresses of the LIT101 and T101 contracts, which can subsequently be utilized by the client application for interaction purposes. This architectural aspect is depicted in Listing 14.

```

//SPDX-License-Identifier: UNLICENSED
pragma solidity ^0.8.9;

import "../LevelSensor.sol";
import "../Tank.sol";

contract ProcessStage {

    address public levelSensorAddr;
    address public tankAddr;

    constructor() {

        Tank tank = new Tank();
        tankAddr = address(tank);
    }
}

```

```

        LevelSensor levelSensor = new LevelSensor();
        levelSensorAddr = address(levelSensor);
    }

    function getChildContractAddresses() public view returns (address
        , address) {

        return (tankAddr, levelSensorAddr);
    }
}

```

Listing 14. Smart Contract used for Process stage

6.5.4 Deployment Script

To facilitate the deployment of the smart contracts, the script depicted in Listing 15 is employed. This script features an asynchronous function designed to gather the contract, specifically the ProcessStage contract in this scenario, and subsequently deploy it along with its child contracts. Upon deployment, the contract address can be queried on Etherscan Sepolia¹⁶ to verify its successful deployment alongside two additional internal contracts. This verification process can be achieved by accessing the transaction details, as exemplified in Figure 18.

```

async function main() {

    const ProcessStage = await hre.ethers.getContractFactory("
        ProcessStage");

    //deploy the processtage contract which also deploys Tank and
    // Level Sensor contracts
    const processStage = await ProcessStage.deploy();

    // wait for the deployment completion
    await processStage.deployed();

    //Get all the addresses

    const [tankAddr, levelSensorAddr] = await processStage.
        getChildContractAddresses();

    console.log(`Process Stage deployed to:  ${processStage.address}`);
    console.log(`Tank Contract deployed to:  ${tankAddr}`);
}

```

¹⁶<https://sepolia.etherscan.io/>

```

    console.log(`Level Sensor Contract deployed to:  ${levelSensorAddr
    }`);
}

// We recommend this pattern to be able to use async/await everywhere
// and properly handle errors.
main().catch((error) => {
    console.error(error);
    process.exitCode = 1;
});

```

Listing 15. Script used for deploying the contracts to Ethereum

To get more details about the proposed solution’s smart contract component, please refer to this public GitHub repository¹⁷.

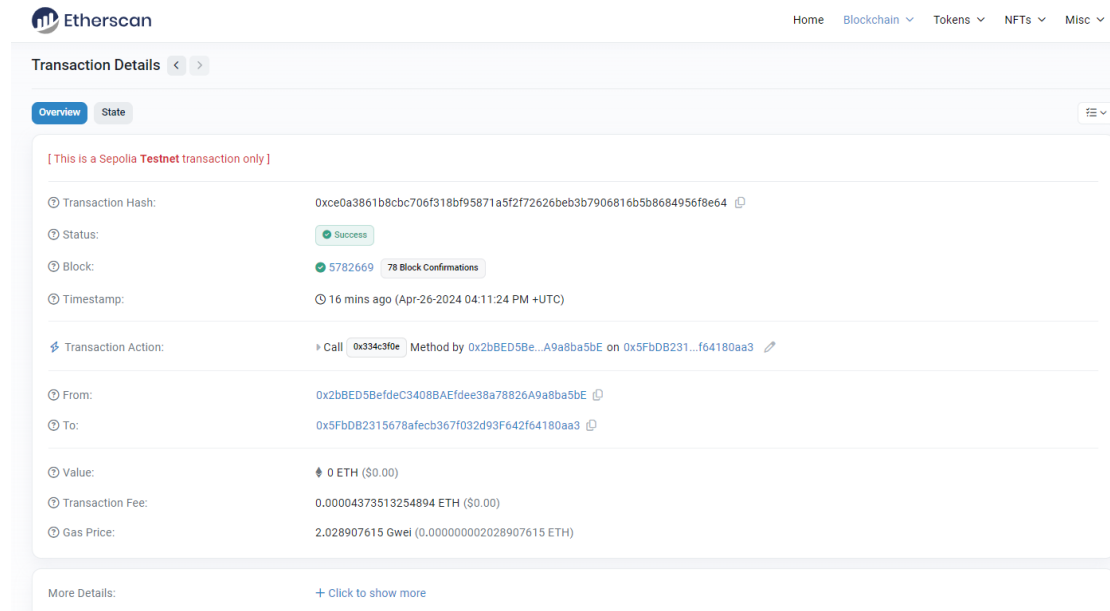


Figure 18. Successful Deployment of contracts showing in Etherscan

6.6 Azure Digital Twin Explorer

A cloud-based ADT Explorer has been used for the proposed P1 stage SWaT architecture (shown in figure 9) simulation to create DTs from models described in Section 6.2.

¹⁷<https://github.com/Ojusvt/thesis-DigitalTwinsAttackDeception/tree/main/smart-contracts>

All JSON-LD-based DTDL files were successfully uploaded to the ADT Explorer¹⁸ application, operating on a cloud server. Subsequently, DTs were instantiated by utilizing the models accessible within the explorer interface. An Excel file defining the requisite relationships and initial property values was uploaded to establish relationships between these twins. Numerous twins corresponding to various devices were generated using the uploaded models, ensuring comprehensive representation within the DT environment. Since a very specific stage process, i.e., the P1 stage process, has been used for creating a DT model and simulation, the created DT has two Flow Sensors, FIT101 and FIT201, Level Sensor LIT101, Tank T101, Pump P101, and Motorized Valve MV101. The structure of the created DT, which is a perfect mimic of the P1 stage SWaT Process, can be seen in figure 19.

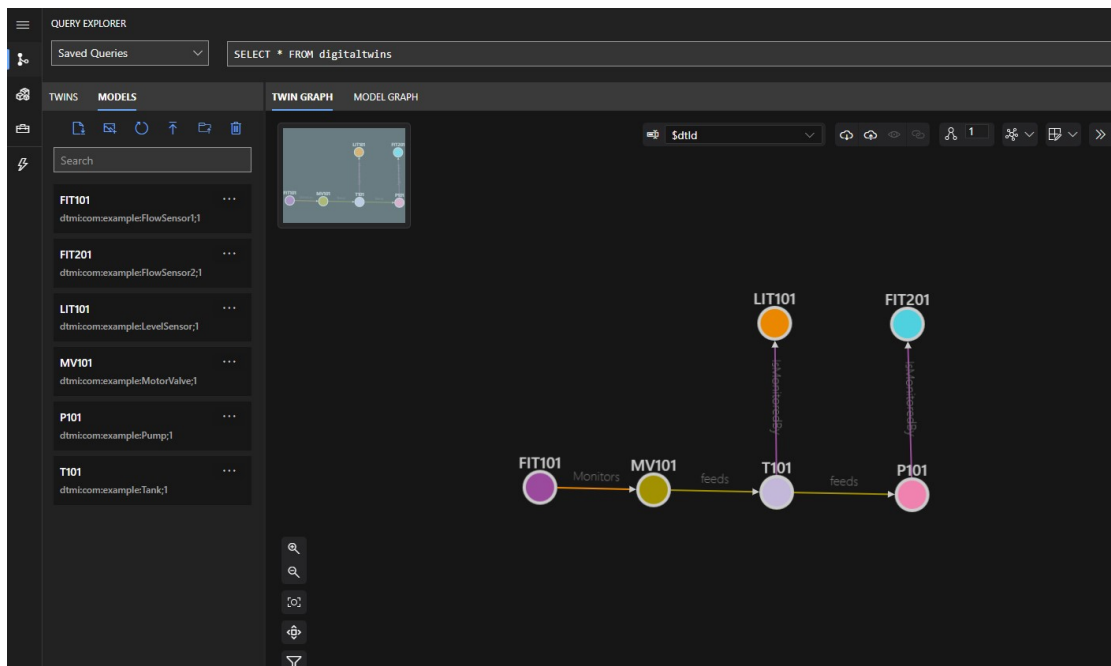


Figure 19. Digital Twins with relationships in ADT Explorer

6.7 Summary

The implementation phase of this study rigorously executed the proposed solution elucidated in Section 5, aiming to enhance the cybersecurity posture of Water CPS through the utilization of DT technology. Beginning with the development of a robust architectural framework integrating DT as an attack deception mechanism, the implementation

¹⁸<https://learn.microsoft.com/en-us/samples/azure-samples/digital-twins-explorer/digital-twins-explorer/>

proceeded to construct, deploy, and simulate the DT for the P1 stage process of the SWaT Testbed. Augmenting the DT's efficacy, an Ethereum-based Solidity Smart contract was employed to safeguard the integrity of telemetry data in the face of insider attacks. This comprehensive approach underscores the pivotal role of DT technology in fortifying the resilience of Water CPS against evolving cybersecurity challenges.

7 Evaluation

This section delineates the assessment of the proposed solution outlined in Section 5, addressing the main research question of the work **RQ: How can Digital Twins be used to Enhance the Cybersecurity of Water CPS?**. The primary purpose of this work was to utilize the DT to enhance the cybersecurity of the Water CPS. In the scope of this work, DT was created and implemented as the attack deception mechanism. The architectural framework and the design of implementing the DT as an attack deception mechanism are provided in Section 6.1. In the later part of this section, the proposed solution is evaluated with the implementation of IRP by considering an attack scenario mentioned in Section 4. The utilization of the proposed solution in the IRP portrays the potential of DT integrated with Blockchain as an attack deception mechanism in the real attack scenario.

7.1 Incident Response Playbook Implementation

This Section illustrates the implementation of IRP based on the attack scenario utilizing the DT integrated with Blockchain as an attack deception mechanism. The flowchart is used to portray the IRP as shown in figure 20. As mentioned in the section 5.6, CISA and NIST guidelines are followed in creating the IRP. The IRP was created using the open-source tool Draw IO ¹⁹. As shown in figure 20, the IRP is created by taking into picture the different roles of the personnel working in the Water CPS. The blocks and the actions mentioned in the blue dotted lines are for the role of CPS Operator, and the blocks and actions mentioned in the red dotted lines are for the role of cybersecurity personnel. The arrows in the IRP are numbered from 1 to 15, which denote the steps to be taken in order to contain the attack in the preliminary phase and effectively use the DT as an attack deception mechanism. The color coding of the blocks is used to show the phases of IRP following the NIST guidelines, such as Detection, Containment, Response, Remediation, Analysis, and Post-incident Activity(explained in detail in section 5.6).

In case of any attacks on the P1 stage process mentioned in table 6, the CPS Operator and Cybersecurity personnel can follow the steps mentioned below to contain and eradicate the attack while keeping the attacker engaged in the DT to collect Indicators of

¹⁹<https://app.diagrams.net/>

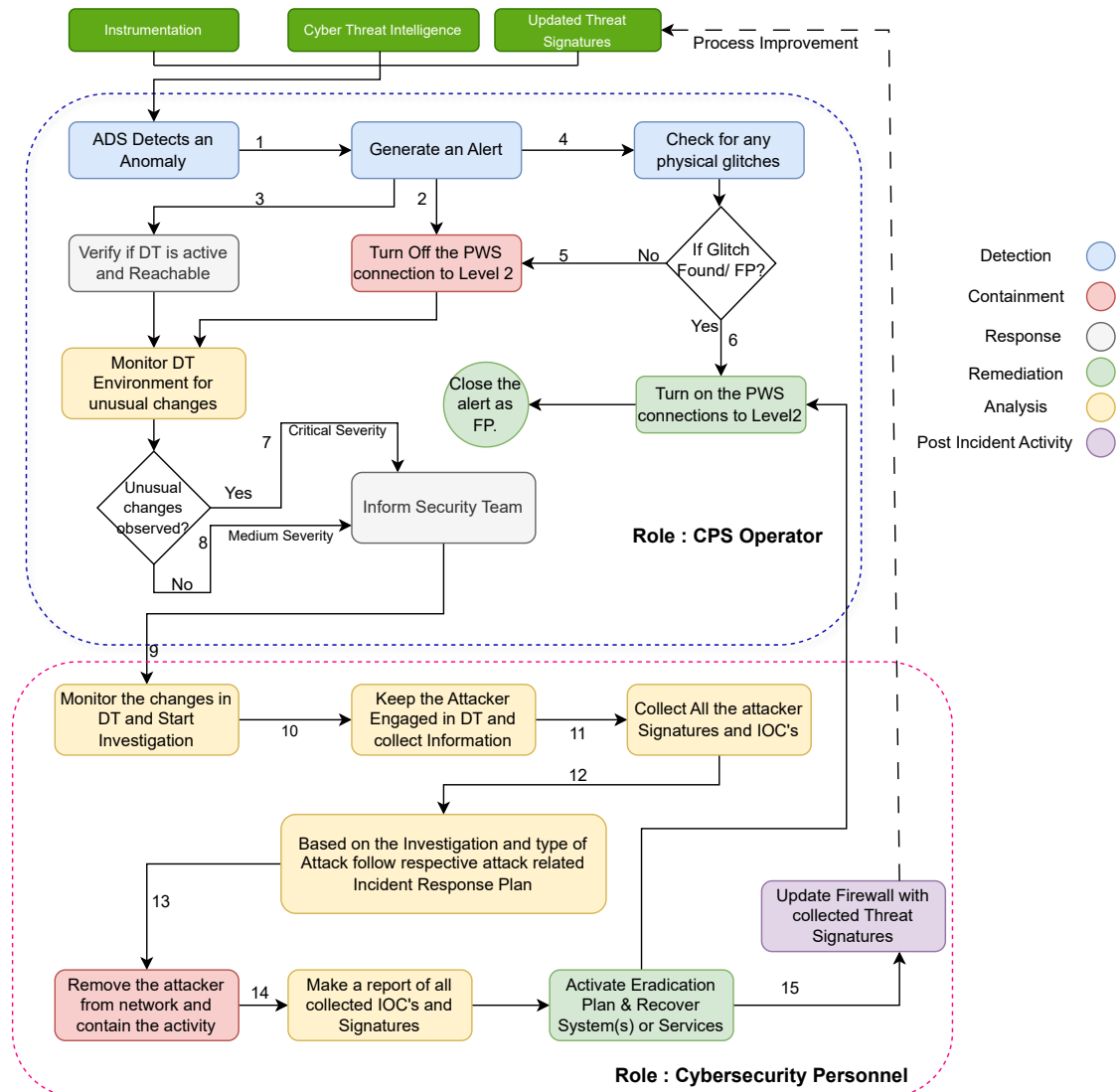


Figure 20. Incident Response Playbook With DT As Attack Deception. Numbers 1-15 denote the Steps.

compromise (IOC) and defend the physical process from further damage or malicious modifications. The steps are explained as a real-time Incident response plan, taking an example attack scenario from the section 4.3 and table 6.

- **Step 1:** If ADS detects the Anomaly, the alert is generated for the CPS operator and Cybersecurity Team.
- **Step 2:** The moment the alert is received. CPS Operator turns off the network connections between Level 3 (SCADA, HMI, Workstations) and Level 2 (Physical

process and respective PLC).

- **Step 3:** While the Network connections are turned off between Level 2 and Level 3, the CPS operator verifies if the DT, strategically placed in the network between Level 2 and Level 3 connected through SW1 with open ports, is active and reachable.
- **Step 4:** As there are multiple possibilities of false positives related to physical issues in the process, which can create an anomaly, the CPS Operator checks for physical glitches in the physical components, which, in fact, will take time. Thus, the network connections from Level 3 are still kept OFF, and the response process can be moved to Step 7 of the IRP.
- **Step 5:** If the physical glitch was not found by the CPS Operator, the connections to Level 3 are kept OFF, and the DT environment must be monitored for unusual modifications in the property values of DT Components.
- **Step 6:** If some physical glitch is found in some components, the CPS operator turns ON the connections between Level 3 and Level 2, and the alert is closed as a false positive.
- **Step 7:** If multiple unusual changes in the DT Environment, which looks like attack numbers 21,35,26 and 30 from table6 are observed, the alert is marked as a critical severity alert and informed to the security team.
- **Step 8:** If no unusual changes are observed in DT except for the behavioral anomaly detected, the alert is marked as a medium severity alert and is communicated to the security team.
- **Step 9:** Cybersecurity Personnel start the investigation by monitoring the changes in DT, which is possible by using multiple security solutions; one example would be Microsoft Defender for IOT²⁰.
- **Step 10:** As the DT is implemented as an attack deception mechanism, the attacker engages in DT and makes real-time changes, assuming the DT is a real physical component. Meanwhile, cybersecurity personnel collect information on the changes made and attacker data.
- **Step 11:** Cybersecurity Personnel collect all the signatures and IOC's and perform in-depth analysis using various threat intelligence tools.

²⁰<https://www.microsoft.com/en-us/security/business/endpoint-security/microsoft-defender-iot>

- **Step 12:** Based on the in-depth analysis and investigation, the type of the attack can be recognized, and the respective attack-related incident response plan can be followed, which varies for every security product.
- **Step 13:** Based on the investigation, collected IOCs, and following the incident response plan, turn off the DT and network and contain the activity. The procedures for this vary for every security vendor.
- **Step 14:** Make a report of all the collected IOCs and signatures, activate the Eradication plan, and recover systems and services. Once the Eradication and recovery are complete, Turn on the network connection between Levels 2 and 3.
- **Step 15:** Update the Firewall with the collected threat signatures and feed the data and learned behavior for process improvement.

The steps mentioned in the IRP are simple and specific to avoid confusion and increase the efficiency of CPS operators and cyber security personnel in case of a real attack scenario. The steps are to be followed in conjunction with the IRP flowchart 20. The detailed explanation of each step considering an attack scenario (discussed in Section 4.3 is as follows:

- **Step 1:** The ADS is deployed in the level 2 of the proposed architecture. The primary aim of this work is to investigate how the DT can be utilized as an attack deception mechanism; thus, various behavior-based detection algorithms proposed (discussed in Section 3 and 5.5) are utilized in this IRP. Once ADS detects an anomaly based on the behavior of the CPS or DT, an alert is generated to cybersecurity personnel on the security tool and CPS operator on any device, such as a phone or dedicated monitor.
- **Step 2:** Once the CPS Operator receives the alert, the key action to be taken is to turn off the connection of the physical process PLC to Level 2 components such as SCADA and HMI. This ensures the incident is contained to prevent the CPS from further damage. This can be just turning off the switch SW2 or might involve a complex process based on CPS's specific architecture. In this step, the assumption is the SCADA in Levels 2 and 3 is compromised as the attacker is able to manipulate the real physical process.
- **Step 3:** Once the actions in step 2 are taken, the CPS Operator must verify if the DT is active and reachable. The DT is deployed in the cloud, and verifying the status of the DT plays an important role. As the DT is placed in close proximity to the real physical twin and is connected to the physical twin with a fake broken connection with open ports to lure the attacker, the manipulation of the real physical twin is mimicked in the DT. DT is constructed similarly to the real physical twin,

understanding the behavior of a real physical twin with fake properties. Thus, attackers are likely to land in the DT environment with open ports, considering it a physical process. But if this is not the case and attacker does the hard job and reach the physical twin and raise an alert. Once the connections are turned off, the attacker would land automatically in DT as it is the only component present in the network.

- **Step 4:** As there are high chances of false positives based on the physical glitches in the physical process, the CPS operator should check for the same while the connections to level 2 are kept off. As this might take time to check for glitches, meanwhile the alert is generated for the concerned security team.
- **Step 5:** If there is no luck finding the physical glitch, it is safe to assume that some external factor is responsible for the anomaly. Thus, the CPS Operator monitors the DT environment for unusual changes in the property values of DT Components.
- **Step 6:** If the CPS operators find a physical glitch that caused the anomaly, the connections to Level 2 are turned on, the normal operations are resumed, and the alert is closed as false positive. This step plays an important role in reducing the downtime of the CPS.
- **Step 7 and Step 8:** While the CPS Operator monitors the DT environment if multiple unusual changes in the DT Environment are observed, which looks like attack numbers 21,35,26 and 30 from table6 are observed, the alert is marked as a critical severity alert and informed to the security team. If no unusual changes are observed in DT except for the behavioral anomaly detected, the alert is marked as a medium severity alert and is communicated to the security team. The Severity level often defines the scale of the attack in cybersecurity terms, making it easy to manage the workload and time of the security team.
- **Step 9:** The security team thoroughly monitors the DT and starts the investigation using the security tool. One good example to investigate the Azure DT alert is the Microsoft Defender for IOT.
- **Step 10:** The changes happening in the DT prove that the attacker is trying to manipulate the property of DT components, thinking of it as a physical device. DT provides real-time simulation, which tricks the attacker into thinking of it as a physical device (explained in detail in table 1). Meanwhile, the security team collects multiple attacker-related information and lets the attacker manipulate the DT.
- **Steps 11 and 12:** Once the required attack data is collected, such as IP addresses, the hashes of multiple executables run, which are called signatures and IOCs in cybersecurity terms, the related incident response plan is followed.

- **Step 13:** This action plan is crucial as the attacker is being removed from the network to contain the activity. This may involve multiple steps based on the type of attack; one example would be to turn off the DT environment, restart the services, and Deploy the DT with a different network configuration.
- **Step 14:** The security team makes a detailed report of all the collected IOCs and signatures, activates the Eradication Plan based on the set standards, and recovers the systems and services back to normal operations. The Eradication and recovery plans may vary for different organizations, security products, and attack types.
- **Step 15:** The collected IOCs and signatures are further fed to the firewall database and various threat intelligence tools so that if something similar occurs again, the firewall blocks the connections outside of the network.

In the construction of IRP, it is made sure to keep it general as much as possible to increase the compatibility and applicability across various Water CPS and cybersecurity products. The Incident Response Strategy needs process improvement from time to time. The proposed IRP is a novel approach based on which the detailed IRP can be constructed for different Water CPS specifically. The steps mentioned in the IRP considering an attack scenario showcase the potential of DT integrated with Blockchain as an efficient attack deception mechanism in various attack scenarios.

7.2 Summary

The step-by-step creation, implementation, simulation of DT of the P1 stage process of SWaT and evaluation with the IRP are provided in Sections 6.2,6.4,6.3 and 7.1, which proves the potential of DT to act as an attack deception mechanism in the attack scenarios mentioned in the table 6 and Section 4.3.2. To fortify the Data integrity of telemetry data supplied to the DT during the insider attack scenarios detailed in section 4.3, an Ethereum-based Solidity Smart contract was Implemented and deployed, which is explained in Section 6.5. This integration of Ethereum-based Solidity Smart contract with DT augments security measures, enhancing the efficacy of DT as an attack deception mechanism. The implementation of the customized role-based IRP leveraging DT as an attack deception mechanism in Water CPS (elucidated in the section 7.1) furnishes vital responses to queries such as *"What incident response procedures can be followed to contain the attack in case of the attack scenario?"* This query assumes significance as revealed by the Literature Review (Section 3 and Table 4), where the absence of robust incident response strategies was identified as a primary factor behind the success of numerous cyberattacks on Water CPS over time. Furthermore, this section answered the question **RQ4: How can an Incident Response Playbook be created to enhance the incident response strategy in Water CPS?**. The customization and deployment of

role-based IRPs, harnessing DT as an attack deception mechanism, reinforced incident response capabilities against potential cyber threats.

8 Discussion

The discussion section of this work delves into the multifaceted realm of DTs, particularly within the context of the Water CPS. Beyond conventional 3D modeling, DTs serve as dynamic entities that seamlessly integrate real-time data and facilitate bidirectional communication with physical entities within virtual environments. In this work, DTs are harnessed as an innovative approach to bolster cybersecurity measures within Water CPS. By leveraging DTs as attack deception mechanisms, potential threats can be identified and mitigated proactively, thereby enhancing the resilience of water infrastructure against cyberattacks. Furthermore, DTs integrated with Blockchain play a pivotal role in ensuring data integrity and traceability, thereby fortifying security measures. Their versatile nature not only enables the detection and prevention of malicious activities but also fosters collaborative efforts and promotes standardization within the realm of water CPS security. This work also addresses the context of Water CPS, SWaT, and the importance of securing the same from potential threats. The lack of extensive research in the Security of Water CPS is highlighted, and the potential of DTs integrated with Blockchain and IRP in enhancing the cybersecurity of Water CPS is discussed. The preceding sections have elucidated the design, implementation, and evaluation of a comprehensive framework aimed at bolstering the cybersecurity of Water Cyber-Physical Systems (CPS) through the innovative integration of DTs. This discussion section delves into the answers to research questions, limitations, and future directions of the research endeavor.

8.1 Answers to Research Questions

The work aimed to address the overarching research question "**How can Digital Twins be used to Enhance the Cybersecurity of Water CPS?**" by investigating four specific sub-research questions. In this section, We discuss the outcomes and results of the work in the context of our research questions. The mapping of research questions to their corresponding study sections is summarized in Table 7.

Firstly, the **RQ1:What role can Digital Twin play in improving the security posture of Water CPS?** focuses on the understanding role of DT in improving the security of Water CPS, which is elaborately discovered from the Section 3 and explained in the section 3.4.2. In this section, we conducted SLR following the Kitchenmann guidelines [21], providing a foundational reference for our proposed solution. This investigation delved into the architectural frameworks facilitating DT integration into Water CPS security, providing insights into the potential mechanisms for bolstering system resilience across the sections 6.1, Section 5.1 providing the answer to **RQ2:What**

Table 7. Mapping of Research Questions with Work

RQs	Description	Section addressing RQ
RQ1	What role can Digital Twin play in improving the security posture of Water CPS?	Section 3.4.2
RQ2	What architectural frameworks facilitate the integration of Digital Twins into Water CPS security?	Section 5.1, 6.1
RQ3	How can Blockchain technology enhance the security of Digital Twins in Water CPS?	Section 5.3, 6.5
RQ4	How can an Incident Response Playbook be created to enhance the incident response strategy in Water CPS?	Section 7.1

architectural frameworks facilitate the integration of Digital Twins into Water CPS security?. This section plays a crucial role in achieving the thesis’s objective of presenting a DT integrated with Blockchain as an attack deception mechanism to enhance the security of Water CPS. Secondly, the study explored the pivotal role of blockchain technology in enhancing the security of DTs within Water CPS, providing the answer to the **RQ3:How can Blockchain technology enhance the security of Digital Twins in Water CPS?** in the section 5.3. This Section 5.3 shed light on the utilization of Ethereum-based Solidity Smart contracts fortifying the data integrity, ensuring the reliability of telemetry data in the face of insider attacks. In Section 6, We implemented the proposed solution using Azure DT services and Ethereum blockchain to further answer the questions **RQ2** and **RQ3** in detail.

Furthermore, the development and implementation of role-based IRPs underscore the importance of proactive incident management strategies in securing critical infrastructure. By delineating clear response protocols, roles, and responsibilities, IRPs empower stakeholders to contain, mitigate, and recover from cyber incidents effectively. This proactive approach mitigates potential damages and minimizes system downtime, securing water system integrity and public safety. Hence, answering the **RQ4: How can an Incident Response Playbook be created to enhance the incident response strategy in Water CPS?** in Section 7.1.

8.2 Limitations

While the proposed framework demonstrates promising capabilities, it is essential to acknowledge several limitations encountered during the research process. Firstly, scalability emerged as a significant challenge, particularly in terms of the adaptability of the proposed framework across diverse Water CPS architectures. Due to the complexities

inherent in real-world systems, the simulated implementations may not fully capture the intricacies and nuances of actual cyber threats and system dynamics, thus limiting the general applicability of the findings. The DT relies on accurate system design information to simulate behavior effectively. However, obtaining precise design data can be challenging, leading to uncertainties in the model's accuracy. This limitation highlights the need for early integration of the DT in the design stages of its physical counterpart to ensure seamless adaptation of design knowledge [10]. The cost spent on implementing and maintaining the DT-based attack deception mechanism with a dedicated cybersecurity team can be quite high for small water CPS.

Furthermore, resource constraints, including time, funding, and computing power, posed considerable challenges throughout the implementation phase. These limitations highlight the necessity for further research and development efforts to address the complexities of securing Water CPS against evolving cyber threats. Additionally, it is important to note that the solution could not be tested in a real Water CPS environment due to various constraints, and the proposed solution does not contain any details regarding network-related implementation as it may vary for every Water CPS. Limited permissions and access prevented the implementation and testing of the proposed framework in an actual operational setting. Moreover, the scope of the research was confined to the P1 stage process of the SWaT architecture. This decision was made due to the complexity of implementing DTs for the entire SWaT architecture, which would have required extensive time and resources. Several technical limitations were also encountered during the implementation phase. For instance, the simulated implementation was conducted solely on the Azure platform, utilizing IoT Hub and ADT services. While this approach provided valuable insights, implementing the solution on other platforms and addressing various scenarios required additional time and resources. Furthermore, limitations associated with integrating DTs with Blockchain, such as complex channels and membership services, impacted the scalability and robustness of the solution.

Moreover, challenges related to developing and implementing an effective Incident Response Playbook for Water CPS cybersecurity were encountered. Creating tailored playbooks that adequately address the unique challenges of securing Water CPS systems requires careful consideration of incident detection, containment, and recovery strategies. Future research endeavors should aim to overcome these challenges and further explore the capabilities of DTs integrated with Blockchain in enhancing cybersecurity measures in Water CPS environments.

8.3 Future Work

Despite the advancements made in this research, several avenues for future exploration and improvement remain. Firstly, addressing the scalability challenges of the proposed framework across diverse Water CPS architectures warrants further investigation. Future research endeavors should enhance the adaptability and scalability of DTs integrated

with Blockchain to accommodate different system architectures and configurations. In the context of cybersecurity, it is imperative to acknowledge the potential weaknesses of the DT itself. Attackers could, for example, trick the DT into thinking that the physical system is safe when it is actually under attack by feeding it consistent historical data [10]. This emphasizes how crucial it is to have secure communication between the physical system and the DT to stop manipulations of this kind [10].

Additionally, there is a need to conduct real-world testing and validation of the proposed framework in operational Water CPS environments. Obtaining permission and access to conduct experiments in real-world settings would provide invaluable insights into the effectiveness and practicality of the solution. Furthermore, expanding the scope of the research beyond the P1 stage process of the SWaT architecture to encompass the entire system would be a fruitful area for future exploration. Technical enhancements and optimizations are also essential for improving the robustness and efficiency of the proposed solution. Exploring alternative platforms and technologies for implementing DTs integrated with Blockchain could yield novel insights and address existing limitations. Additionally, further research is needed to refine and optimize IRPs tailored for Water CPS cybersecurity, considering the evolving nature of cyber threats and attack vectors. Moreover, extending the research to the next processes of the SWaT architecture is imperative to provide a comprehensive cybersecurity solution for Water CPS. Investigating the implementation of DTs and incident response strategies in subsequent processes beyond the P1 stage would offer a holistic approach to securing critical water infrastructure.

Lastly, collaboration with industry stakeholders and regulatory bodies is crucial for ensuring the practicality and applicability of the proposed solutions in real-world settings. Establishing partnerships with water utility companies and cybersecurity experts would facilitate the implementation and adoption of cybersecurity measures in Water CPS environments, ultimately enhancing the security and resilience of critical water infrastructure.

9 Conclusion

This work presents a novel and comprehensive approach to enhancing the security of Water Cyber-Physical Systems through the implementation of a Digital Twin integrated with Blockchain as an attack deception mechanism. By leveraging DTs and blockchain technology, we address the pressing need for robust cybersecurity measures in critical infrastructure.

Throughout this thesis, we conducted a systematic literature review to explore the landscape of cybersecurity threats and detection mechanisms targeting water CPS. The review revealed many security threats and vulnerabilities inherent in water systems, underlining the critical importance of robust cybersecurity measures to safeguard against

cyber-physical attacks. Notably, the review identified the potential of advanced technologies such as AI, ML, and DTs in enhancing the resilience and security of water infrastructure. In response to the findings of the SLR, we developed a comprehensive solution framework that integrates DTs with Blockchain technology. Our framework not only addresses existing vulnerabilities but also emphasizes the importance of proactive incident response strategies. By developing IRPs tailored to Water CPS environments, we provide a road map for effective threat containment and mitigation. Through meticulous design and implementation, we have demonstrated the efficacy of our approach in mitigating cyber threats and enhancing the security posture of water CPS.

Nevertheless, our research encountered inherent limitations and challenges, including scalability issues, resource constraints, and the complexity of real-world implementations. Despite these challenges, our work represents a significant step forward in bolstering the cybersecurity of water infrastructure.

In conclusion, our proposed system offers a promising avenue for enhancing the security of Water CPS. By bridging the gap between digital and physical systems and emphasizing proactive incident response strategies, we pave the way for a more resilient and secure water infrastructure that can withstand the ever-evolving threat landscape. As we continue to refine and iterate upon our framework, we remain committed to safeguarding critical infrastructure and ensuring the integrity of water systems worldwide.

References

- [1] Openai chat. <https://chat.openai.com/>. Accessed: April 2024.
- [2] Emmanuel Aboah Boateng, J. W. Bruce, and Douglas A. Talbert. Anomaly detection for a water treatment system based on one-class neural network. *IEEE Access*, 10:115179–115191, 2022.
- [3] Ahmed Abokifa, Kelsey Haddad, Cynthia Lo, and Pratim Biswas. Detection of cyber physical attacks on water distribution systems via principal component analysis and artificial neural networks. pages 676–691, 05 2017.
- [4] Ahmed Abokifa, Kelsey Haddad, Cynthia Lo, and Pratim Biswas. Real-time identification of cyber-physical attacks on water distribution systems via machine learning based anomaly detection techniques. *Journal of Water Resources Planning and Management*, 145, 07 2018.
- [5] Hajar Addeen, Yang Xiao, Jiacheng Li, and Mohsen Guizani. A survey of cyber-physical attacks and detection methods in smart water distribution systems. *IEEE Access*, PP:1–1, 07 2021.

- [6] Hajar Hameed Addeen, Yang Xiao, Jiacheng Li, and Mohsen Guizani. A survey of cyber-physical attacks and detection methods in smart water distribution systems. *IEEE Access*, 9:99905–99921, 2021.
- [7] Sridhar Adepu and Aditya Mathur. Using process invariants to detect cyber attacks on a water treatment system. volume 471, 05 2016.
- [8] Sridhar Adepu, Jay Prakash, and Aditya Mathur. Waterjam: An experimental case study of jamming attacks on a water treatment system. In *2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pages 341–347, 2017.
- [9] F. Batubara, Jolien Ubacht, and Marijn Janssen. Challenges of blockchain technology adoption for e-government: a systematic literature review. pages 1–9, 05 2018.
- [10] Mehmet Bozdal. Security through digital twin-based intrusion detection: A swat dataset analysis. pages 1–6, 10 2023.
- [11] F-Secure. Havex Hunts For ICS/SCADA Systems. <https://www.f-secure.com/weblog/archives/00002718.html>. Accessed on 3 February 2024.
- [12] Nicolas Falliere, Liam O. Murchu, and Eric Chien. W32.Stuxnet Dossier (Version 1.4). White paper, Symantec Security Response, Mountain View, CA, USA, 2011.
- [13] Marcio Giacomoni, Nikolaos Gatsis, and Ahmad Taha. Identification of cyber attacks on water distribution systems by unveiling low-dimensionality in the sensory data. pages 660–675, 05 2017.
- [14] Jonathan Goh, Sridhar Adepu, Khurum Junejo, and Aditya Mathur. A dataset to support research in the design of secure water treatment systems. 10 2016.
- [15] Jonathan Goh, Sridhar Adepu, Marcus Tan, and Zi Shan Lee. Anomaly detection in cyber physical systems using recurrent neural networks. In *2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE)*, pages 140–145, 2017.
- [16] Michael Grieves. Digital twin: Manufacturing excellence through virtual factory replication. 03 2015.
- [17] Amin Hassanzadeh, Amin Rasekh, Stefano Galelli, Mohsen Aghashahi, Riccardo Taormina, Avi Ostfeld, and M. Katherine Banks. A review of cybersecurity incidents in the water sector. *Journal of Environmental Engineering*, 146(5), May 2020.

- [18] Mashor Housh and Ziv Ohar. Model-based approach for cyber-physical attack detection in water distribution systems. *Water Research*, 139:132–143, 2018.
- [19] Jun Inoue, Yoriyuki Yamagata, Yuqi Chen, Christopher M. Poskitt, and Jun Sun. Anomaly detection for a water treatment system using unsupervised machine learning. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 1058–1065, 2017.
- [20] Kaspersky. BlackEnergy APT Attacks in Ukraine. <https://www.kaspersky.co.uk/resource-center/threats/blackenergy>. Accessed on 3 February 2024.
- [21] Barbara Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering. 2, 01 2007.
- [22] Nir Kshetri. Blockchain’s roles in strengthening cybersecurity and protecting privacy. *Telecommunications Policy*, 09 2017.
- [23] Min Kyoung Won, Young Hwan Choi, and Joong Kim. Development of a cyberattack detection model for a water distribution system using water quality and hydraulic criteria. *Journal of the Korean Society of Hazard Mitigation*, 20:47–56, 10 2020.
- [24] Lupovis. The ultimate guide to cyber deception technology. <https://www.lupovis.io/the-ultimate-guide-to-cyber-deception-technology/>. Accessed on October 10, 2023.
- [25] Aditya P. Mathur and Nils Ole Tippenhauer. Swat: a water treatment testbed for research and training on ics security. In *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*, pages 31–36, 2016.
- [26] Satoshi Nakamoto and A Bitcoin. Bitcoin: A peer-to-peer electronic cash system. <https://bitcoin.org/bitcoin.pdf>, 2008.
- [27] NTT Ltd. The rapid evolution of deception technologies. <https://services.global.ntt/-/media/ntt/global/solutions/intelligent-cybersecurity/secure-infrastructure/the-rapid-evolution-of-deception-technologies---whitepaper.pdf>. Accessed on October 10, 2023.
- [28] M. Pasha, Bijay KC, and Saravanakumar Somasundaram. An approach to detect the cyber-physical attack on water distribution system. pages 703–711, 05 2017.
- [29] Morgen Peck. Blockchains: How they work and why they’ll change the world. *IEEE Spectrum*, 54:26–35, 10 2017.

- [30] Jacopo Pisani, Graziana Cavone, Federica Pascucci, and Laura Giarre. Using digital twin to detect cyber-attacks in industrial control systems. In *IEEE EUROCON 2023 - 20th International Conference on Smart Technologies*, pages 467–471, July 2023.
- [31] Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. In Eric Goetz and Sujeet Sheno, editors, *Critical Infrastructure Protection*, pages 73–82, Boston, MA, 2008. Springer US.
- [32] Sabah Suhail, Rasheed Hussain, Raja Jurdak, Alma Oracevic, Khaled Salah, and Choong Seon Hong. Blockchain-based digital twins: Research trends, issues, and future challenges. *CoRR*, abs/2103.11585, 2021.
- [33] Symantec. W32.Duqu: The Precursor to the Next Stuxnet (Version 1.4). White paper, Symantec Security Response, Mountain View, CA, USA, 2011.
- [34] Riccardo Taormina, Stefano Galelli, Nils Ole Tippenhauer, Elad Salomons, and Avi Ostfeld. Characterizing cyber-physical attacks on water distribution systems. *Journal of Water Resources Planning and Management*, 143, 02 2017.
- [35] Riccardo Taormina, Stefano Galelli, Nils Ole Tippenhauer, Elad Salomons, Avi Ostfeld, Demetrios Eliades, Mohsen Aghashahi, S Asce, Raanju Sundararajan, Mohsen Pourahmadi, M Banks, Bruno Brentan, Enrique Campbell, G Lima, Daniel Manzi, David Ayala-Cabrera, Manuel Herrera, Idel Montalvo, Joaquín Izquierdo, and Ziv Ohar. The battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks. *Journal of Water Resources Planning and Management*, 144, 03 2018.
- [36] TechRepublic. Pennsylvania water system hack demonstrates lax security, 2006.
- [37] TechTarget. How to create an incident response playbook. *SearchSecurity*.
- [38] Trellix. What is soc?, Accessed 2024.
- [39] Tufin. Building an effective soc playbook, Accessed 2023.
- [40] Nilufer Tuptuk, Peter Hazell, Jeremy Watson, and Stephen Hailes. A systematic review of the state of cyber-security in water systems. *Water*, 13, 01 2021.

Appendix

I. Resources

Code Repository: <https://github.com/Ojusvt/thesis-DigitalTwinasAttackDeception.git>

The GitHub repository contains the implementation code of Digital Twin models in DTDL language, Data ingestors, and Simulators of the DT models in CSharp and Solidity Smart contracts. The repository is public.

Demo Videos:

Data Simulators for a proposed solution - Simulators Update Digital twin in Azure Digital Twin explorer : <https://youtu.be/a0flp3gle7s>

The video showcases the execution and working of the simulator code of the Level Sensor LIT101 DT with fake telemetry data and the Updation of the LIT101 DT in the ADT explorer cloud, confirming the successful connection deployment of LIT101 DT and Simulator code.

Data Ingestors for a proposed solution - Interaction with blockchain and Azure Digital Twin explorer : <https://youtu.be/o-OVBq7rUqw>

The video showcases the execution and working of the ingestor code and smart contract of the Level Sensor LIT101 DT with simulator data provided. The updation of the logs for the function app is demonstrated to verify the smart contract analysis and generation of the warning log.

II. Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Ojus Virendra Tudavekar**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

Blockchain and Digital Twin-based Approach for Securing Water Supply Infrastructure,

supervised by Dr. Mubashar Iqbal.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.
4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Ojus Virendra Tudavekar

15/05/2024