High-fidelity cyber and physical simulation of water distribution systems. II Enabling cyber-physical attack localization

Murillo, Andrés; Taormina, Riccardo; Tippenhauer, Nils Ole; Galelli, Stefano

DOI
10.1061/JWRMD5.WRENG-5854

Publication date
2023

Document Version
Final published version

Published in
Journal of Water Resources Planning and Management

ASCE

High-Fidelity Cyber and Physical Simulation of Water Distribution Systems. II: Enabling Cyber-Physical Attack Localization

Andrés Murillo[1]; Riccardo Taormina[2]; Nils Ole Tippenhauer[3]; and Stefano Galelli, M.ASCE[4]

Abstract: A fundamental problem in the realm of cyber-physical security of smart water networks is attack detection, a key step towards designing adequate countermeasures. This task is typically carried out by algorithms that analyze time series of process data. However, the nature of the data available to develop these algorithms limits their capabilities: by relying on process data only, one cannot distinguish a cyber-attack from the failure of a system's component or identify the root cause of an attack. Here, we show that these limitations can be addressed through the joint analysis of process and network data—with the latter representing the information exchanged between the components constituting the Industrial Control System, such as sensors and Programmable Logic Controllers (PLCs). For this purpose, we utilize a dataset generated by digital hydraulic simulator (DHALSIM)—a numerical modelling platform built on a two-way interaction between EPANET version 2.2 and a network emulation tool—which is extended here to include a framework for launching cyber-physical attacks. This paper presents a dataset with realistic network information of a smart water network under cyber-physical attacks and presents an analysis of how that information can enable the development of better intrusion detection systems that can localize and identify attacks. Through this analysis, the dataset provided here, and the open-source availability of DHALSIM, our work paves the way to a novel class of analytics for actionable detection. **DOI: 10.1061/JWRMD5.WRENG-5854.**

Author keywords: Water distribution systems; Smart urban water networks; Cyber-physical attacks; Cyber security; EPANET.

Introduction

The vulnerability of water distribution systems to cyber-physical attacks is an unintended consequence of the progressive digitalization of the urban water sector: the widespread deployment of information and communication technologies allows utilities to monitor and control, in (near) real-time, their entire value chain (Makropoulos and Savić 2019), but also creates a digital 'attack surface' that could be exploited by hackers (Rasekh et al. 2016). The sense of urgency associated with this emerging threat is well demonstrated by the increase in the frequency of cyber-physical attacks, as well as their diversity and complexity (Hassanzadeh et al. 2020). In turn, this calls for a holistic approach to cybersecurity, one that builds on a balanced combination of policy measures (Shapira et al. 2021) and analytics tasked with the problem of identifying vulnerabilities, disclosing cyber-attacks, and designing reactive measures (Tuptuk et al. 2021).

So far, the water community has focused on two main types of analytics: numerical simulation models and intrusion detection systems. The former are a fundamental cornerstone (Berglund et al. 2020): by simulating the physical response of a water distribution system to cyber-threats, we can support a multitude of tasks, such as predicting the likely impact of attacks (Taormina et al. 2017) or identifying the most unfavorable operational conditions (Shin et al. 2020). Currently, there are two models able to support the aforementioned analyses, epanetCPA (Taormina et al. 2016, 2019) and RISKNOUGHT (Nikolopoulos et al. 2020; Nikolopoulos and Makropoulos 2021). The two models differ for various implementation details, but both rely on the same modelling concept, that is, the combination of a hydraulic model [EPANET Rossman (2000)] with a simplified representation of the Industrial Control System's components—i.e., sensors, actuators, Programmable Logic Controllers (PLCs), and Supervisory Control and Data Acquisition (SCADA) system. Simulation models also play a key role in the development of intrusion detection systems, a particularly active research domain [see the recent review by Addeen et al. (2021)]. Because long time series of observational data featuring cyber-attacks are unavailable (Hassanzadeh et al. 2020), detection systems largely rely on simulated datasets, such as the BATADAL (Taormina et al. 2018).

Notwithstanding this recent progress, more methodological advances are needed to design detection algorithms that can fully characterize a cyber-physical attack, thereby leading to 'actionable detection'—using an expression introduced by Tuptuk et al. (2021). A major challenge is the fact that the current class of intrusion detection systems cannot distinguish a cyber-physical attack from an anomaly caused by the failure of a system's component, such as a

[1]Postdoctoral Research Fellow, iTrust Centre for Research in Cyber Security, Singapore Univ. of Technology and Design, 8 Somapah Rd., Singapore 487372 (corresponding author). ORCID: https://orcid.org/0000-0001-6965-2283. Email: andres_murillo@sutd.edu.sg

[2]Assistant Professor, Faculty of Civil Engineering and Geosciences, Delft Univ. of Technology, Stevinweg 1, Delft 2628 CN, Netherlands.

[3]Professor, CISPA Helmholtz Center for Information Security, Stuhlsatzenhaus 5, Saarbrücken 66123, Germany. ORCID: https://orcid.org/0000-0001-8424-2602

[4]Professor, Pillar of Engineering Systems and Design, Singapore Univ. of Technology and Design, 8 Somapah Rd., Singapore 487372. ORCID: https://orcid.org/0000-0003-2316-3243. Email: stefano_galelli@sutd.edu.sg

Note. This manuscript was submitted on May 16, 2022; approved on December 3, 2022; published online on February 22, 2023. Discussion period open until July 22, 2023; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, © ASCE, ISSN 0733-9496.

malfunctioning pump or a defective communication link between two PLCs (Taormina et al. 2018; Ahmed et al. 2020). In addition, detection algorithms struggle to pinpoint the system components that are under attack (Taormina and Galelli 2018). That means we cannot identify the root cause of cyber-physical attacks, since the same hydraulic response can be obtained by completely different attack vectors (Taormina et al. 2017). Naturally, then, we cannot react to an attack if we do not know where it originated—or whether it is an attack at all. The explanation behind these limitations lies in the nature of the data used to train detection algorithms (Addeen et al. 2021): the datasets currently available are generated with models that simulate only hydraulic processes and are based on a simplistic representation of the networked infrastructure. But if we do not also study—and model—the processes occurring within the Industrial Control System, we cannot explain the true nature of an attack.

Here, we address this knowledge gap and lay the foundations for a novel class of detection systems based on the joint analysis of process and network data. In particular, we show that the availability of high-resolution cyber and physical data enables us to pinpoint the root cause of cyber-attacks and other observed anomalies. For our analysis, we leverage a dataset generated by digital hydraulic simulator (DHALSIM)—a numerical modelling platform built on a two-way interaction between EPANET and a network emulation tool (Murillo et al. 2022)—whose functionalities are expanded here to include a framework for launching cyber-security experiments as well as automated scripts for processing the network traffic data. Our analysis is complemented by an introduction to the most common vulnerabilities affecting the industrial communication networks commonly adopted in automated water distribution systems.

Background

We begin by providing background information on Industrial Control Systems (ICSs), industrial communication networks, and network vulnerabilities. As we shall see, it is necessary to understand the relationship between an industrial communication network and a physical system being controlled by an ICS. Readers can gain more in depth detail of ICS and industrial communication networks from the background section of the companion paper (Murillo et al. 2022).

Industrial Control Systems

ICSs are cyber-physical systems designed to guarantee that a physical process operates at all times based on a set of defined operational parameters (Humayed et al. 2017). ICSs are controlled by the integration of a computing platform and an industrial communication network. The computing platform of an ICS is represented by industrial computers called Programmable Logic Computers (PLCs) and a SCADA server. The PLCs perform the following operations: (1) measure the processes occurring in the physical system; (2) exchange that information with other PLCs; (3) make control decisions to maintain the physical process within configured parameters; and (4) apply the control decisions through the actuators. The SCADA server is used to centralize the information exchanged by the PLCs and send configuration parameters to the PLCs. This process is repeated periodically and is known as a *scan cycle*.

Smart water networks are ICSs that automatically maintain hydraulic and water quality parameters within pre-configured values. Fig. 1 shows a simple smart water network. In the network, the level of Tank 1 (T1) is controlled by Pump 1 (P1), while two PLCs (PLC1 and PLC2), one SCADA server, and the industrial communication network make up the ICS. PLC1 uses a sensor to measure the water level of T1; then, it sends this reading to PLC2, which uses the reading to operate P1 (e.g., to decide whether P1 needs to be turned on or off). This scan cycle is executed periodically to maintain the tank level within the desired operational parameters. In addition, the PLCs report the values of different variables to the SCADA server, such as T1 level, P1 status, P1 flow, or the pressure at the junctions. Also, note that each PLC is located in a different substation. This means that each PLC is located within a Local Area Network (LAN) and both networks are connected using a Wide Area Network (WAN), represented in Fig. 1 by r0. Considering that water networks are typically distributed across vast spatial domains, locating the PLCs in different substations is therefore a compelling network configuration. This is because PLCs and their local network need to be allocated near the actuators or sensors that they are controlling, hence a single substation is not enough for a water distribution system.

Industrial Communication Networks

Network communications are logically divided into layers. Each layer has certain functionalities that can be offered by a specific network protocol (Tanenbaum and Wetherall 2010). In this way, protocols are designed to operate at a specific layer and for a specific application; for example, IEEE 802.11b/g/n (commonly known as "WiFi") is a network link layer protocol for wireless networks. Another example is IEEE 802.3 (Ethernet), another network link layer, but designed to be used on wired networks. In addition, each protocol only provides a fraction of the complete functionality required to have a complete communication service, so different layers (with different protocols) are stacked together to offer complete network applications. The term stack is used because protocols at lower layers offer services to the upper layer, and the uppermost layer is the one directly offering the final application.

An Industrial Communication Network (ICN) is a special type of communication network used by PLCs and SCADA servers to exchange information and enable them to control physical processes (Galloway and Hancke 2013). The difference between ICNs and traditional communication networks lies in the protocols used in ICNs. ICNs use industrial communication protocols to exchange information between communication nodes. Two common protocols used are *Ethernet/IP* (ENIP) and *Common Industrial Protocol* (CIP). The latter is an application layer protocol that enables communications between PLCs. Using CIP, PLCs exchange messages containing sensor readings (i.e., tank water levels) or actuator status (i.e., pump status). CIP is supported by ENIP, a protocol that offers a communication session in which CIP messages are exchanged.

The use of ICNs to exchange information to control the physical system creates a relationship between ICNs behavior and the state of the physical system. As a consequence, faults or attacks on different network nodes or links are likely to affect the physical system in a particular manner (Sánchez et al. 2019). This is important, because if enough information about the physical system (e.g., water network architecture, control strategies) and network behavior is available (e.g., network topologies, protocols used), anomalies in the physical system behavior can be traced back to network attacks, and vice-versa. In other words, high-quality information about the behavior of the physical and cyber layers could enable intrusion detection mechanisms that not only detect attacks, but are also able to identify the nature and location within the network of such attacks. This is one of the main reasons motivating the need for modeling tools that combine an accurate representation of physical processes with a realistic implementation of network behavior.

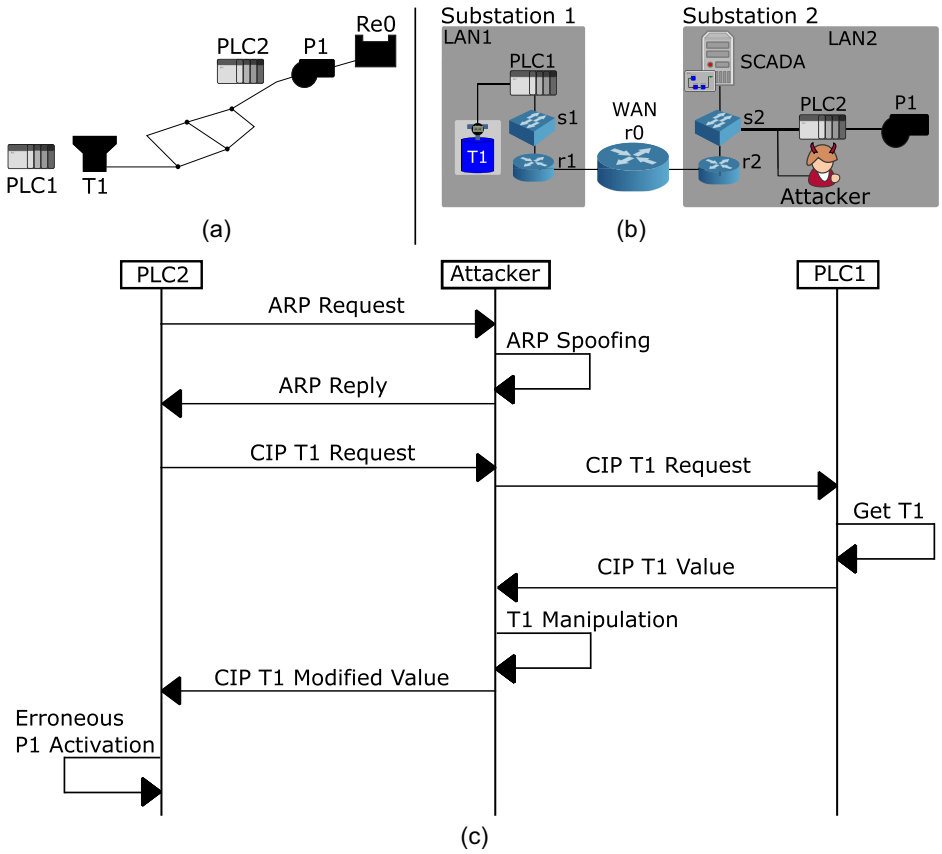

Fig. 1. (a) Illustration of a water distribution system; (b) its industrial control system; and (c) the sequence of network messages exchanged when a PLC obtains a sensor value from another PLC. The water level of Tank T1 is monitored by PLC1, which then relays the information to PLC2. The latter finally controls Pump P1. The water source, or reservoir, is denoted as Re0. In the cyber layer, the two PLCs are located in their own substation. This choice is explained by the spatially-distributed nature of water distribution systems. The PLCs are connected by a set of routers and switches. Note the presence of an attacker located in Substation 2. (c) The network messages and interactions between the PLCs and an attacker performing an ARP spoofing attack to manipulate the reading of Tank T1. After the ARP spoofing, the attacker can intercept all messages of PLC2: when PLC1 sends the T1 tank reading, the attacker manipulates the reading to make PLC2 erroneously activate P1.

Network Vulnerabilities and Attacks in Industrial Communication Networks

A communication network vulnerability is a condition, or property, of a protocol stack that creates an opportunity for launching a cyber attack and compromising one security property of the network. For example, Denial of Service Attacks (DoS) compromise the availability property of a network, making some network nodes (and their information) unreachable to other points of the network. Another example is Man-in-the-Middle Attacks (MiTM), which compromise the integrity of a network by allowing attackers to manipulate the information exchanged in the network.

As an example of a network vulnerability and cyber attack, we analyze a MiTM attack and then its impact in the context of smart water networks. One way to implement an MiTM attack is to exploit one vulnerability of the Address Resolution Protocol (ARP). By way of background, the ARP protocol allows network nodes to learn the addresses of their neighboring devices and the gateways to external networks. In Fig. 1, the ARP protocol could allow PLC2 to know the address of the SCADA server and the address of its gateway, r2. The vulnerability in this protocol is that ARP protocol messages do not use any authentication or integrity protection schemes. This would enable an attack to use *ARP Spoofing* to carry out a MiTM attack. *ARP Spoofing* is a technique used by an attacker to impersonate the gateway of a network node, and this allows the attacker to potentially manipulate all the messages being sent and received by the target. In the use case shown in Fig. 1, an attacker located in LAN2 could use ARP spoofing to attack PLC2, impersonating r2. Then, the attacker could manipulate the level of T1 sent by PLC1 and received by PLC2, causing PLC2 to operate P1 incorrectly (Urbina et al. 2016). Suppose that the attacker only manipulates the readings reaching PLC2 and not PLC1: in this case, the attack would generate information on the physical state that might lead to the identification of the attack (P1 would be operated incorrectly, according to the T1 level reported by PLC1, which is unaffected by the attack). In addition, the network information might reveal the nature of the attack through a set of malicious ARP messages sent between a network node and r2.

Importantly, unexpected behaviors of the physical processes are not necessarily due to cyber-attacks. Two important causes of anomalous behaviors are network events and device attacks. The former are non-malicious conditions that happen in a network, such as a delay in the packets being exchanged or the loss of a given fraction of packets. These conditions might be caused by hardware or software failure at the networking equipment or by network congestion. The latter cause an industrial device to behave incorrectly—for example by activating a pump at an incorrect moment or ignoring the configured control rules. In industrial networks, device attacks can also be the result of devices infected by malware (Sandaruwan et al. 2013).

Digital Hydraulic Simulator

Background

Fig. 2 shows a simplified version of DHALSIM architecture. The configuration files are parsed to create a virtual network topology in Mininet, which is a platform for creating virtual networks running inside a single host machine (Lantz et al. 2010). In DHALSIM, each of the Mininet nodes runs a script representing the PLCs, SCADA, and the water distribution system. The latter node runs an EPANET simulation in a step-by-step fashion, where the duration of each step is the 'hydraulic time-step' (configured in EPANET) and the number of steps run is used as the master clock for the simulation. In addition, if the user has configured network events (e.g., delays in the communication network) or cyber-physical attacks, additional scripts are launched to execute them. These are easily configurable using optional files that are explained below. Finally, the file generator module generates documentation for the events or attacks launched during the simulation and includes the documentation in the result files. Here, we have provided a detailed explanation of the modules used to launch events and attacks and illustrate how to configure experiments with device attacks, network attacks, or network events. Additional details about DHALSIM are provided in (Murillo et al. 2022).

Framework for Launching Cyber-Security Experiments

DHALSIM is a co-simulation environment that uses a distributed approach to run experiments. With this approach, some experiments can include running multiple pieces of code in different Mininet nodes. In order to keep DHALSIM user-friendly and the attacks it simulates configurable, we opted for a framework that requires launching a simple command, instead of multiple commands across multiple nodes. This is achieved using optional configuration files that specify the parameters and types of attacks or events to be launched.

All configuration files, except the EPANET input file, use YAML (YAML Ain't Markup Language). All experiments require three mandatory files: experiment configuration file, EPANET input file, PLCs configuration file. The experiment configuration file defines the global options for a DHALSIM experiment; the EPANET input file is a "standard" EPANET .inp file; and the PLC configuration file indicates how many PLCs are present in the smart water network and which variable they handle. Additional information regarding these files can be found in (Murillo et al. 2022) and in the website documentation: https://github.com/afmurillo/DHALSIM/blob/master/doc/configuration.rst. To launch network attacks or events, the following optional files must be provided:

- Attacks Configuration file: Attacks are configured in this optional file. Attacks provide triggers that can be used to easily set conditions that launch the attack. There are two types of triggers: time triggers and value triggers. The former are activated when the simulation master clock reaches a user specified value. The latter triggers are activated when one sensor in the simulation reaches the user-specified value. Using a value trigger, a device attack could be triggered when a specific tank level reaches the configured value. Two types of attacks are configured in DHALSIM: device attacks and network attacks. Device attacks are attacks running in PLC processes that can change the way a PLC applies a control logic. Network attacks launch an additional Mininet node running a script that exploits a network vulnerability and affects the network and physical behavior. These network attacks are activated by triggers. Attack scripts that implement Denial-of-Service attacks and MiTM attacks are provided in DHALSIM.
- Events Configuration file: Events are configured in this optional file. Currently, only network events are supported. Network events are events that affect the way a network link behaves. An example of an event would be one that causes a percentage of packets in a network link to be dropped. Events also use triggers to launch their execution. Network events do not launch additional Mininet nodes; instead, they run in the routers already present in the Mininet network.

These configuration files are processed by the Parser, and additional scripts might be launched to execute the attacks or events configured by the user. Fig. 3 shows the framework using these files. Every event or attack launched requires a trigger to control when the execution is started and finished. Triggers are based on either physical variables or time. In the first case, DHALSIM starts an attack if the actual value of a variable in the EPANET simulation is below, above, or within a configured threshold. In the second case, the attack starts when the experiment master clock reaches the configured value. The module is the name of the attack or event to be launched. DHALSIM provides a repository with some attacks and events already configured and provides modules that can be used by the community to create additional attacks or events. All attacks and events modules extend a "synced_attack" or "synced_event" module that offers all basic functionalities required by these module. Finally, both attacks and events allow certain options or parameters to be used during execution. For example, for a network event of type "delay," the delay in seconds and the network link must be configured; meanwhile, for a MiTM attack, the target PLC, the tag being manipulated, and the new tag value must be configured.

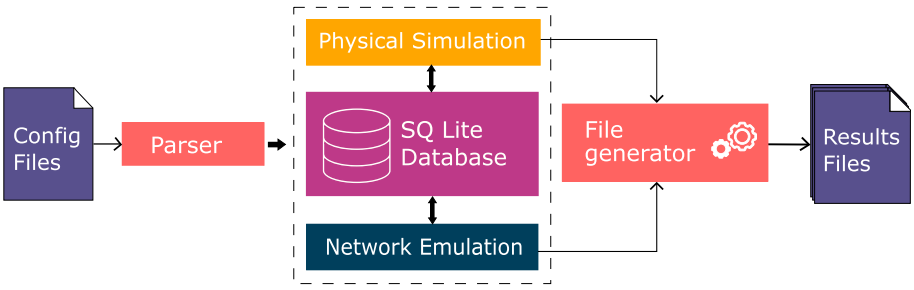

Fig. 2. DHALSIM Architecture. The architecture is composed of a parser, physical simulation, network emulation, an SQLite Database, and a File generator.

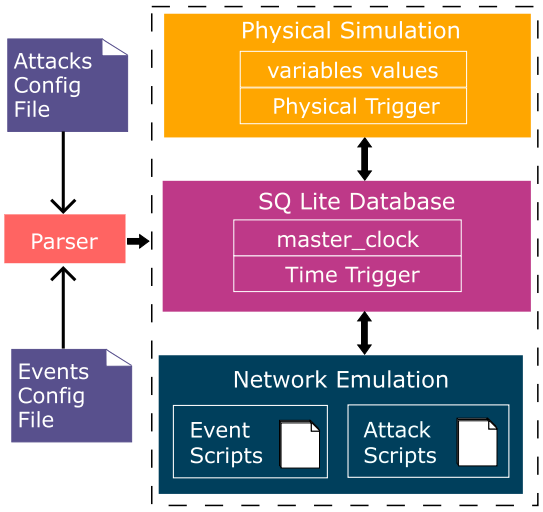

Fig. 3. DHALSIM Attack and Events Framework. The framework allows researchers to include optional attack or events configurations files to run these attacks or events. Attacks and events execute scripts that can be triggered by different conditions such as the simulation clock or values of the physical variables.

Experimental Setup

Case Study

For our analysis we use the case study of Anytown, which is adopted in Murillo et al. (2022) to illustrate the key functionalities of DHALSIM and to study the value of network and process data under a broad range of normal operating conditions. Here, we complement that analysis by unveiling the information that the same type of data can provide for a system undergoing cyber-physical attacks. As shown in Fig. 4, the physical layer of Anytown consists of one reservoir, two tanks (T41 and T42), and two pumps (P78 and P79) controlling the water level of the tanks. The monitoring and control process relies on three PLCs. Specifically, PLC2 and PLC3 monitor the water level of the two tanks and send this information to PLC1, which operates the pumps. Both pumps follow the same control rule and are turned on when the water level in the tanks drops below 5 m (the status of pump P78 is a function of tank T41, while the one of pump P79 depends on tank T42). The SCADA server and the three PLCs belong to separate substations (and corresponding local area networks), thereby reflecting the spatially distributed nature of the water distribution system.

All experiments are run for one week, using a hydraulic time-step of five minutes and pressure-driven analysis (Douglas et al. 2019). We use the same initial tank levels and demand pattern for all the experiments outlined below. Finally, the experiments are carried out on an Intel Xeon (R) 81 W-2175 CPU 2.5 GHz with 128 GB of RAM running Linux Ubuntu 20.04 (Focal Fossa). With this hardware, the run time for a single one-week simulation with DHALSIM is about 24 min.

Normal Operating Conditions

In this experiment, all components of the cyber-physical system (e.g., pumps, sensors, communication links) work in normal operating conditions. For this scenario, we run DHALSIM over a span of seven days, starting on Monday at 00:00 and finishing at Sunday at 23:59. We use this experiment to establish a baseline for the system behavior and to illustrate the interactions between communication network and physical system.

Threat Model

We consider an attacker that can compromise the communication link between PLC1 and r2. We assume that the attacker is familiar with the smart water network and knows the control strategy used to control the pumps. In addition, the attacker is able to parse and modify the network messages of CIP/ENIP sent through the

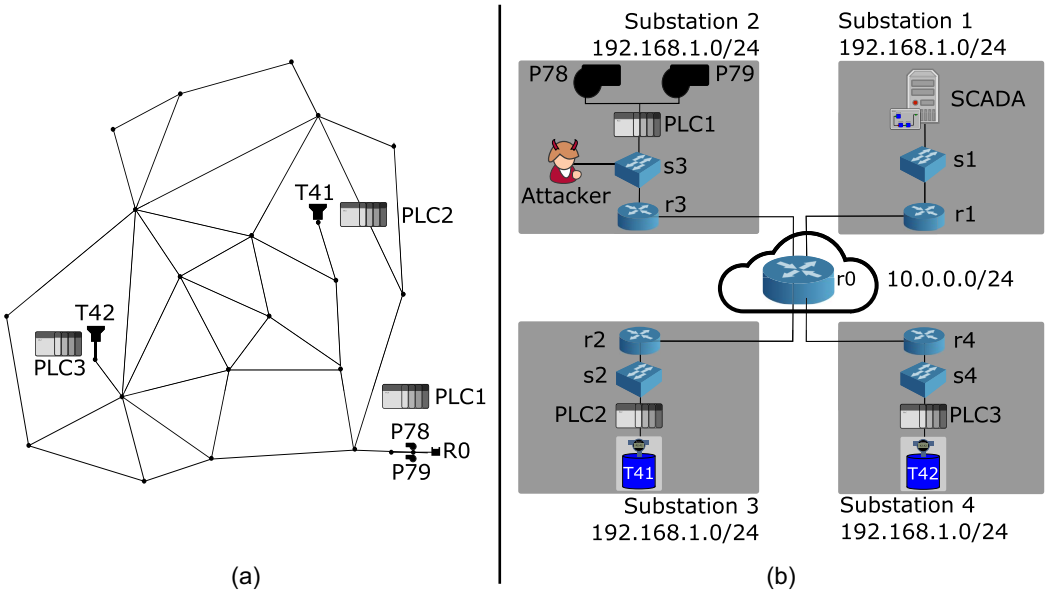

Fig. 4. (a) Physical layer; and (b) cyber-layer of the Anytown water distribution system. The physical layer consists of two tanks (T41, T42) controlled by pumps P78 and P79 (a). Three PLCs control the system: PLC2 and PLC3 monitor the water level of the two tanks and send this information to PLC1, which operates the pumps. Each PLC belongs to a substation. Note the presence of an additional substation, where the SCADA server is located. During the cyber attacks, the attacker is located in Substation 2 (b).

Table 1. Scenarios deployed. All scenarios target PLC1 and start and end at the same time. The first two scenarios are cyber-attacks while the third one is a network connectivity issue. Note that all scenarios are run using the same initial tank levels and demand patterns

Scenario	Type	Cause	Starting time	Ending time	Target
Scenario 1	Attack	ARP Spoofing DoS	Wednesday, 06:00	Wednesday, 18:00	PLC1
Scenario 2	Attack	ARP Spoofing MiTM	Wednesday, 06:00	Wednesday, 18:00	PLC1
Scenario 3	Failure	Linux Traffic Control	Wednesday, 06:00	Wednesday, 18:00	PLC1

network. We also assume that the attacker is not able to compromise the PLCs in the network or the SCADA server.

Anomalous Operating Conditions

We created three modelling scenarios characterized by different types of attacks but with the same hydraulic response. In particular, we model a hypothetical situation in which tanks T41 and T42 are temporarily empty: this is a challenging anomaly to analyze for an operator relying only on SCADA readings, because of the many potential causes behind the anomaly—e.g., cyber-physical attacks or a sensor, pump, or network malfunction. Notice that the kind of analysis presented here is only possible with a simulation tool or a dataset that accurately models the network communication behavior of an industrial control system. The scenarios have the following specifications (see Table 1 for further details):

- Scenario 1 depicts a situation in which an attacker located in Substation 2 [Fig. 5(a)] carries out a DoS attack. The attacker first uses the ARP spoofing method to intercept all readings arriving at PLC1 and then stops forwarding the messages to the PLC, preventing it from receiving new readings on the tank water levels. In turn, this forces PLC1 to make control decisions based on outdated information, eventually leading to low water levels in the tanks (Krotofil et al. 2014).
- Scenario 2 represents a MiTM attack carried out from Substation 2 [Fig. 5(b)]. This attack also relies on ARP spoofing to intercept the messages arriving at PLC1. In this case, however, the attacker manipulates the tank water level readings to make PLC1 believe that the tanks are full. This forces PLC1 to turn pump P78 and P79 off, emptying the tanks.
- Scenario 3 is a network malfunction scenario (i.e., no attack). It represents a situation in which 100% of the packets arriving to PLC1 are temporarily lost due to networking failure. Similarly to Scenario 1, PLC1 is forced to control the pumps using outdated information on the tank levels. In DHALSIM, we implement this scenario using a Linux network tool called tc (traffic control). This tool is integrated in DHALSIM and used by the network events scripts. Moreover, this scenario is similar to some of the network conditions analyzed in the Part 1 of this paper (Murillo et al. 2022).

Although DHALSIM is capable of adding noise to the WDS sensors, as explained in the Part 1 of this paper (Murillo et al. 2022), these experiments do not add noise, because we wanted to make sure that the only impact in the physical results was caused by the cyber attacks and anomalies.

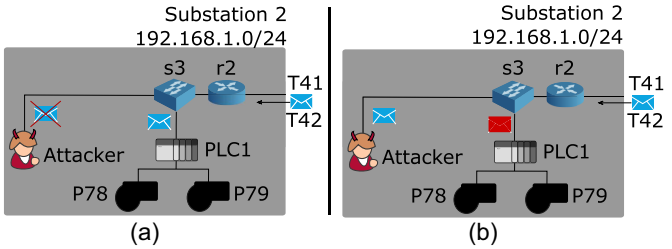

Fig. 5. Attacks in DHALSIM case study: (a) Denial-of-Service Attack; and (b) MiTM Attack. In the Denial-of-Service Attack, the attacker uses an ARP spoofing attack to intercept the messages arriving at PLC1. Then, the attacker stops forwarding those messages to PLC1, causing it to take control decisions with outdated information. In the MiTM attack, the attacker uses an ARP spoofing attack to manipulate the messages arriving at PLC1. The attacker manipulates the water tank levels to cause PLC to take control decisions with incorrect information.

Results

Physical Response

Fig. 6 shows the water level in tanks T41 and T42 during normal operating conditions and under Scenarios 1–3. The results show that the hydraulic response of the water distribution system is rather similar during the attacks and network malfunction. As mentioned above, it would be challenging for an operator to identify the root cause of this problem if relying only on SCADA data—e.g., time series of the tank water levels. Notice that existing simulation tools that only represent the physical system and do not accurately model the industrial communication network would not be able to produce meaningful data to differentiate between these attacks. Thus, the intrusion detectors developed with only physical system data would not be able to localize or identify the nature of the attack or anomaly.

Explaining the Physical Response

To fully characterize the nature of the observed anomalies in the tank water levels, we base our analysis on all data generated by DHALSIM, namely the process data retrieved from the SCADA server and the network packet captures of the PLCs. With the aid of these data, we create a diagnostic flowchart that illustrates the different steps an operator could follow to pinpoint the root cause of the anomaly (Fig. 7). The flowchart is divided in two regions, marked by gray boxes. The first region represents the analysis the operator could perform with SCADA data, and the second region shows the analysis that would require access to network data. Note that epanetCPA (Taormina et al. 2016; Taormina et al. 2019), Risknought (Nikolopoulos et al. 2020; Nikolopoulos and Makropoulos 2021), and DHALSIM could all be used in the first region, but only DHALSIM could be used for the second region, as it is the only tool that provides network data. Also, note that the steps are reported in increasing order of complexity—so only the last ones require to analyze the network packet captures. For the sake of simplicity, we assume that the communication between SCADA and PLCs has not been compromised; another scenario that could be simulated with DHALSIM. The network data are stored in *.pcap* files. These files store all network messages sent and received by one node in pcap format (DHALSIM generates one.pcap file for each PLC and SCADA in an experiment).

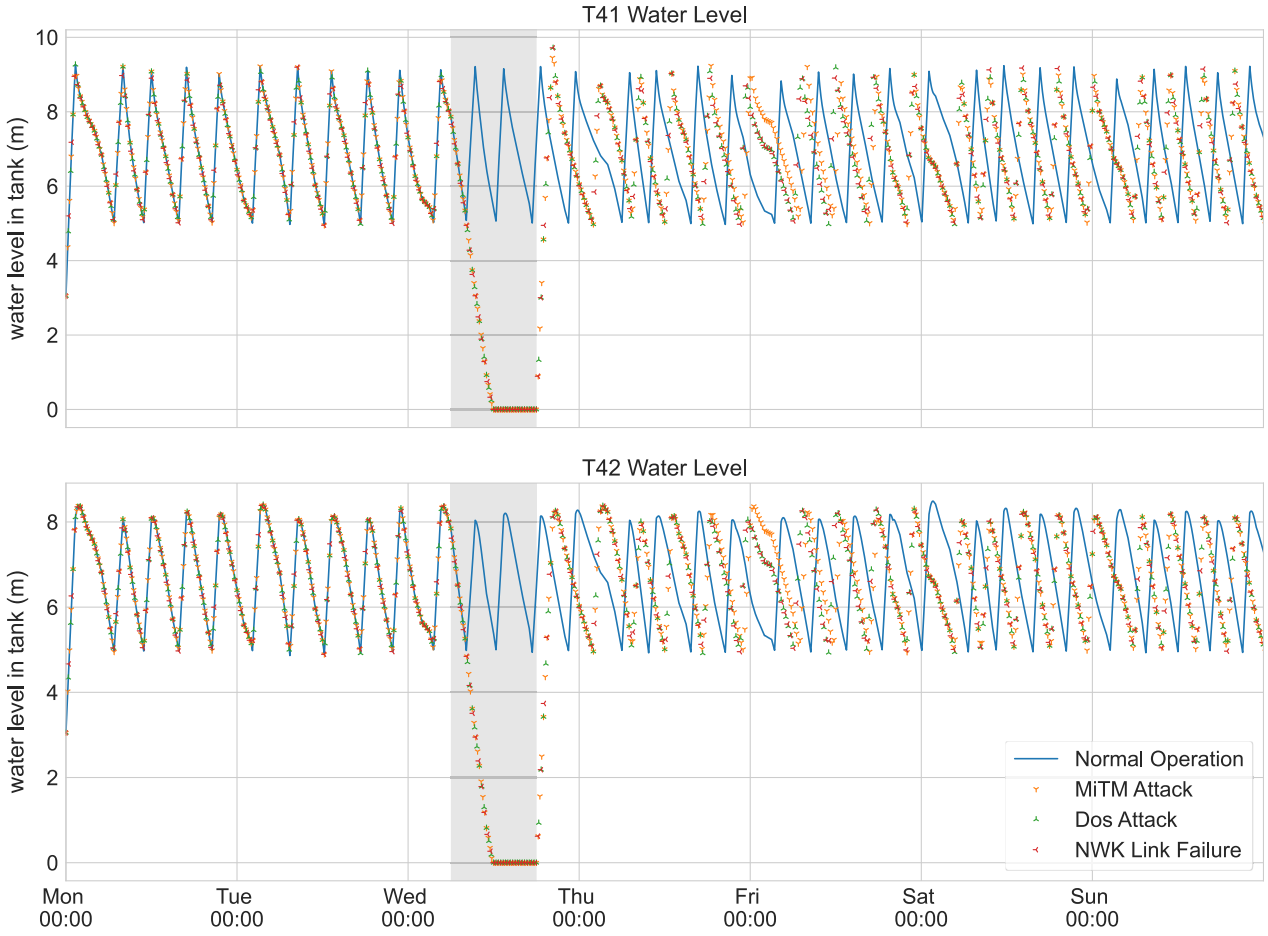

Fig. 6. Physical response of Anytown to different anomalies. Trajectory of the tank water levels under four scenarios—normal operating conditions, DoS attack (Scenario 1), MiTM attack (Scenario 2), and network failure (Scenario 3). Note that the two attacks and network failure cause a very similar response.

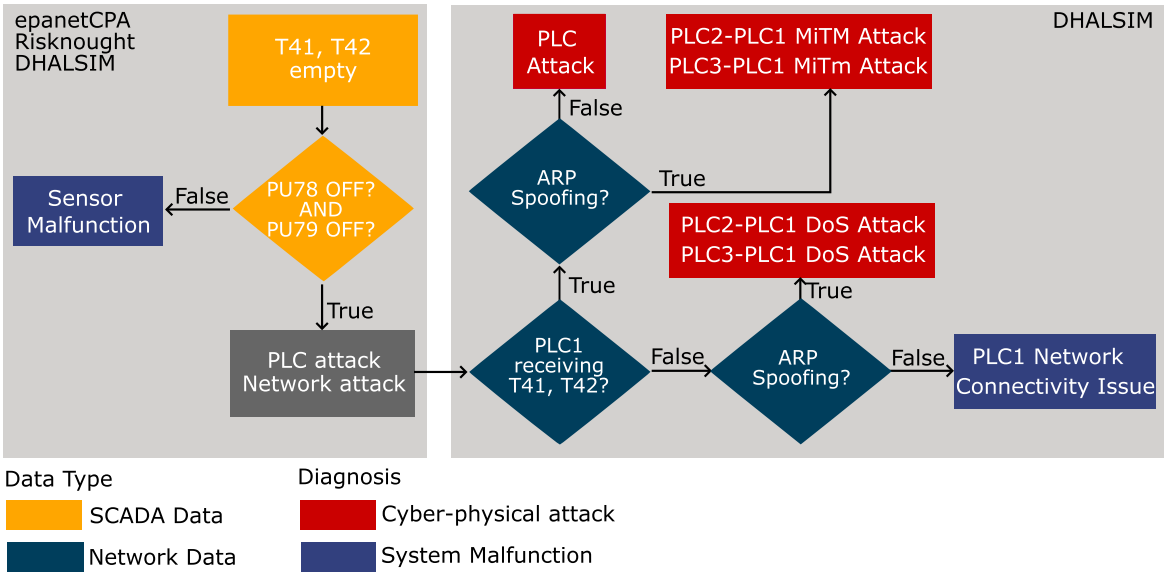

Fig. 7. Diagnosis flowchart for anomalies in Anytown. The flowchart identifies all possible causes leading to the anomalous water level in Tank T41 and T42 illustrated in Fig. 6. This analysis is based on the assumption that the SCADA communication is still secure. The flowchart is divided in two regions. On the left side, we illustrate the first steps of the diagnostic exercise, which could be carried out with the data produced by epanetCPA, Risknought, or DHALSIM. On the right, we illustrate the remaining steps. Note that the latter can be supported only by DHALSIM.

To process these .pcap files, we used a library named *scapy* (Kobayashi et al. 2007) and two specific parsers for ENIP and CIP (Urbina et al. 2016).

As shown in Fig. 7, the first condition an operator could check are the states of pumps P78 and P79. If their state is coherent with the one expressed by the control rules, then the tanks cannot be empty (provided that the water demand is within normal parameters and that there are no sudden infrastructural issues, such as a pipe burst). The observed anomaly could therefore be explained by a sensor malfunction in tank T41 and T42. If none of these two conditions is verified—and so both pumps and sensors are working properly—then the culprit must be the control actions applied by PLC1. Such a situation could be due to three different reasons: first, the PLC is receiving outdated information (Denial-of-Service attack or network malfunction); second, the water level readings received by the PLC have been manipulated (MiTM attack); third, the PLC is malfunctioning or is under direct attack. To rule out each condition, we must now analyze the network packet captures. Notice that at this point, we would have to stop our analysis if we were using epanetCPA or Risknought.

Using DHALSIM, we could continue by analyzing the total number of packets received by PLC1 during all simulations (Fig. 8, upper panel), an indicator of whether PLC1 is receiving outdated information. The drop in packets in Scenario 3 is a good indication that PLC1 might be operating with outdated information due to a network connectivity issue or an attack. On the contrary, the number of packets for Scenario 2 hints that PLC1 might be receiving updated information, but it is possible that this information is being manipulated by an attacker. Finally, the anomalous increase in packets of Scenario 1 could be caused by the network having to resend multiple packets that did not arrived to their destination or by a Denial-of-Service attack aimed at PLC1.

To further diagnose the cause of the scenarios, we analyze the number of reset (RST) packets received by PLC1. These are

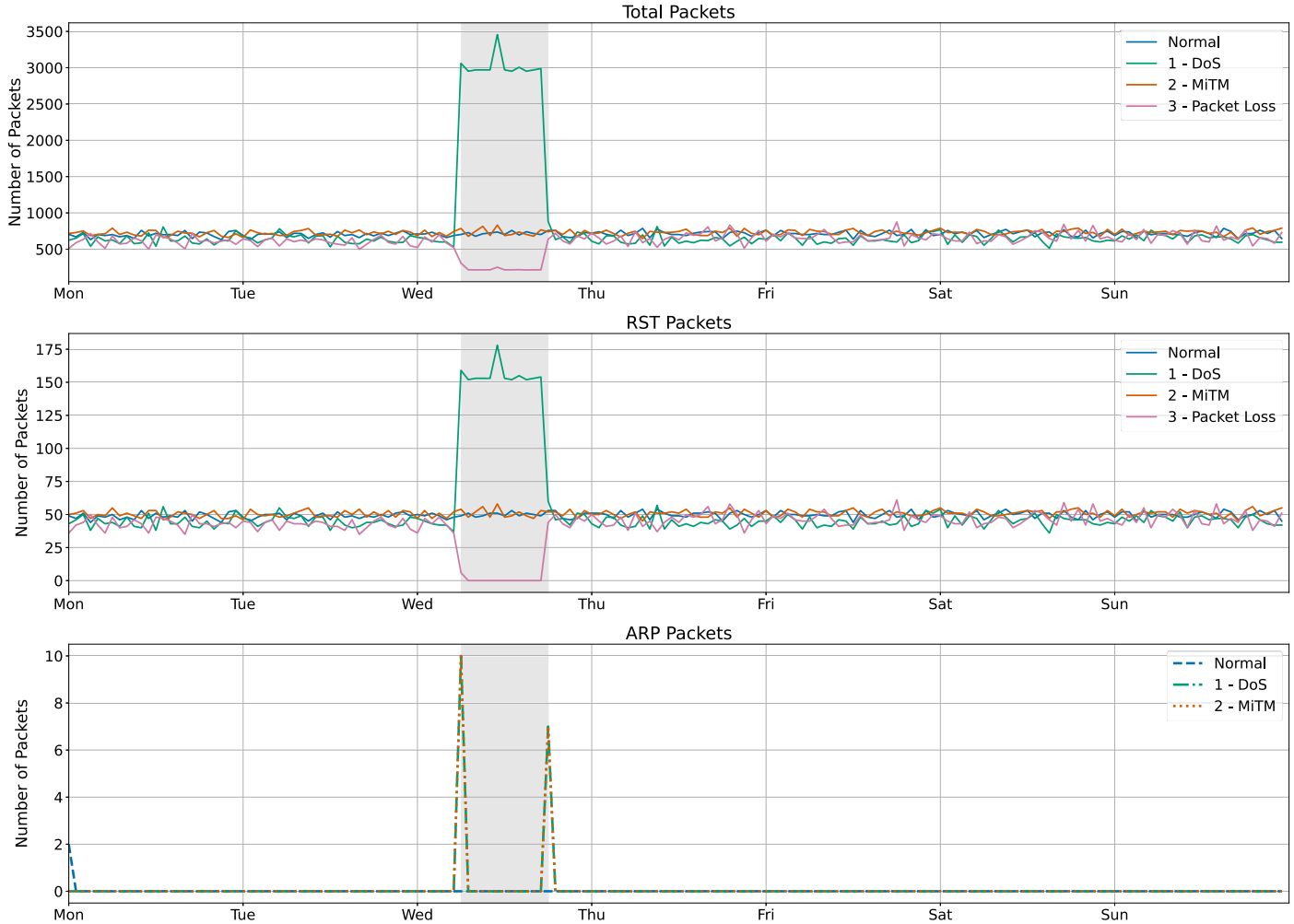

Fig. 8. Packets per hour received by PLC1. In the upper panel, we report the total number of packets received by PLC1. Note that in Scenario 1 (Denial-of-Service attack) the number of packets reaching PLC1 increases heavily. This is because the Denial-of-Service attack generates retransmission of packets and the unexpected ending of TCP connections (RST packets, shown in the middle panel). In Scenario 2 (MiTM), the total number of packets remains the same. This is because the MiTM attack does not affect heavily the number of packets sent; rather, it manipulates the payload (tank readings) of the packets sent in the network. In scenario 3 (Packet Loss), the total number of packets almost reaches zero, because the network connectivity issue completely stops the traffic towards PLC1. In the bottom panel, we illustrate the number of ARP messages received by PLC1. In normal operating conditions, PLC1 should only receive ARP messages at the beginning of the simulation. The spikes present between at the beginning and ending of the anomaly indicate the presence of a possible ARP spoofing attack, which is associated to either a Denial-of-Service (Scenario 1) or an MiTM attack (Scenario 2).

packets sent by a network node when the connection is abruptly finished (which could happen during a DoS attack). Fig. 8 (middle panel) shows that Scenario 1 is the only one that has also an increase in this number of packets. In addition, Scenario 2 has levels similar to normal operating conditions, while Scenario 3 shows a slight drop in the number of RST packets. So far, we have evidence of an anomalous number of packets being received in Scenarios 1 and 3. Moreover, Scenario 3 presents a drop in the number of RST packets, meaning that the anomalous physical response of Scenario 3 is most likely caused by a network connectivity issue.

How to further distinguish between Scenario 1 and Scenario 2? To answer this question, we analyze the number of ARP messages received by PLC1, a good indicator of Denial-of-Service (Scenario 1) and MiTM (Scenario 2) attacks. In normal operating conditions, ARP messages are present only at the beginning of an experiment. This is because, when an experiment begins, the PLCs do not know the Media Access Control (MAC) addresses of their gateways and so ARP messages are exchanged between the PLCs to get the MAC address, which is stored in the ARP cache. (This is an artifact of DHALSIM simulations, as real systems would have already exchanged this information.) Another situation that can trigger the transmission of ARP messages is when nodes are not able to communicate with each other for long periods of time, causing the ARP cache to expire. Notice that both DoS and MiTM require executing an ARP spoofing attack, that is, the process of sending additional ARP messages to poison the ARP cache of a PLC. As shown in Fig. 8 (bottom panel), the signature of the two attacks is well reflected in a small number of ARP messages received by PLC1 right at the beginning and ending of the attack. This means that in both scenarios there was an ARP spoofing attack. However, we also know that in Scenario 2 (MiTM) there was no increase in the total number of packets received by PLC1 [Fig. 8 (upper panel)]. With this, we can conclude that the ARP spoofing of Scenario 2 is a MiTM attack: PLC1 receives roughly the same number of packets with T41 and T42 readings, but these readings are manipulated to cause PLC1 to operate the pumps wrongly. Meanwhile, in Scenario 1 (DoS), the ARP spoofing attack causes PLC1 to drop all the packets with T41 and T42 readings arriving at PLC1, causing an increase in the number of RST packets (Fig. 8, middle panel).

Discussion and Conclusions

Our results show that the joint analysis of process data and network traffic allows us to identify the root cause of cyber-physical attacks and rule out other possible anomalies, such as a sensor malfunction or a temporarily loss of information received by a PLC. This finding has a direct implication on the field of attack detection algorithms, which are currently designed to only analyze process data [see, for example, Taormina and Galelli (2018), Abokifa et al. (2019), Chandy et al. (2019), Ramotsoela et al. (2019), Kadosh et al. (2020), and Tsiami and Makropoulos (2021)]. We can thus envision a novel generation of algorithms that relies on data pertaining to both hydraulic and network traffic, such as the packets per hour illustrated in Fig. 8. In addition, the performance of these algorithms could be tested in real-time on DHALSIM. For example, DHALSIM could be extended to run a Python module with the code of a detection algorithm that sits at the SCADA node to gather real-time data on hydraulic and network processes. The distributed nature of DHALSIM would make such analysis flexible, as one could also deploy multiple detection algorithms in different nodes (corresponding, for instance, to different PLCs) or even conceive a scenario in which a detection mechanism is tested against an adversarial algorithm tasked with the problem of conceiving its attacks (Erba et al. 2020).

Naturally, the co-simulation of hydraulic and network processes offered by DHALSIM lend itself to many other types of applications in the realm of cyber-physical security. For example, this extended version of DHALSIM could be used to study the vulnerability of different network topologies to cyber-physical attacks—that is, evaluating how alternate configurations of sensors and PLCs or different distributions of substations and network topologies connecting them may unintentionally expand the attack surface. Such analysis is particularly important for medium- and large-scale systems, where analysts may have to choose between many possible topologies (Shin et al. 2020). For a given system and number of sensors, for example, a small number of PLCs would limit the investment and operational costs associated to the ICS, but would also require to connect multiple sensors to the same PLC. The latter could therefore become a critical element, since the unavailability of just a single PLC would impede the operators from monitoring or controlling large portions of the water system. On the other hand, increasing the number of PLCs would require to deploy multiple local area networks—as well as network routing to interconnect them—thereby exposing the cyber-physical systems to other vulnerabilities (Abe et al. 2016). Finding a reasonable trade-off between costs, vulnerabilities, and operational flexibility is thus a complex task that could be tackled with DHALSIM. The opportunities are many, from exploring alternate topologies via numerical simulation to coupling DHALSIM with a global search algorithm tasked with the problem of minimizing the vulnerability to cyber-physical attacks.

It is worth noting that DHALSIM could also be used to run advanced network analyses that use different network topologies or networking tools, such as Netcat or Ettercap. These off-the-shelf tools could be used to run network vulnerability analyses or network attacks in a very similar way of how these activities are carried out in the real world. For example, a researcher could use Netcat to discover the protocol being used by the PLCs and then Ettercap to launch an MiTM attack to hijack that communication. Because there are many off-the-shelf tools for network analysis, we preferred not to integrate them directly in DHALSIM, therefore providing flexibility to the user to choose which network tool to integrate—a task supported by DHALSIM's configuration tools and experiment triggers. We conclude with a final note on DHALSIM and cyber security: the protection of any cyber security infrastructure is a multi-layered approach that relies on multiple mechanisms and practices working together to achieve objectives such as integrity, confidentiality, and availability. In the case of traditional IT services, there is a more widespread practice of secure protocols, integrity protection, and access control mechanisms. Nevertheless, the adoption of such practices in ICS faces difficulties. First, not all ICSs can deploy secure protocols because they have legacy hardware that may not support them, or the protocols used have constraints that do not make it possible to use such features. One example of this issue is common field bus protocols (Kayan et al. 2022; Dzung et al. 2005). Second, there is an endemic gap between cyber security good practices and their actual implementation, especially in the water sector (Hassanzadeh et al. 2020), as also demonstrated in the very recent attack on South Staffordshire Water (Labs 2022). As such, we believe it is still important to create tools like DHALSIM that enable modellers to simulate worst-case scenarios, where these secure protocols are not presented or misconfigured, or where an attacker has compromised the identity of authorized nodes in the network. Cyber security does not have a silver bullet; instead, it provides security through multiple layers of defense and tools.

Data Availability Statement

DHALSIM is available at https://github.com/afmurillo/DHALSIM. Some or all data, models, or code generated or used during the study are available in a repository or online in accordance with founder data retention policies. The dataset is available at https://zenodo.org/record/6323248.

Acknowledgments

This research is supported by Singapore's National Satellite Of Excellence, Design Science and Technology for Secure Critical Infrastructure (NSoE DeST-SCI) through the project "LEarning from Network and Process data to secure Water Distribution Systems (LENP-WDS)" (Award No. NSoE_DeST-SCI2019-0003) and by the Faculty of Civil Engineering and Geosciences of Delft University of Technology.

References

Abe, S., M. Fujimoto, S. Horata, Y. Uchida, and T. Mitsunaga. 2016. "Security threats of internet-reachable ICS." In *Proc., 2016 55th Annual Conf. of the Society of Instrument and Control Engineers of Japan (SICE)*, 750–755. New York: IEEE.

Abokifa, A. A., K. Haddad, C. Lo, and P. Biswas. 2019. "Real-time identification of cyber-physical attacks on water distribution systems via machine learning-based anomaly detection techniques." *J. Water Resour. Plann. Manage.* 145 (1): 04018089. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001023.

Addeen, H. H., Y. Xiao, J. Li, and M. Guizani. 2021. "A survey of cyber-physical attacks and detection methods in smart water distribution systems." *IEEE Access* 9: 99905–99921. https://doi.org/10.1109/ACCESS.2021.3095713.

Ahmed, C. M., M. R. Gauthama Raman, and A. P. Mathur. 2020. "Challenges in machine learning based approaches for real-time anomaly detection in industrial control systems." In *Proc., 6th ACM on Cyber-Physical System Security Workshop, CPSS '20*, 23–29. New York: Association for Computing Machinery.

Berglund, E. Z., J. E. Pesantez, A. Rasekh, M. E. Shafiee, L. Sela, and T. Haxton. 2020. "Review of modeling methodologies for managing water distribution security." *J. Water Resour. Plann. Manage.* 146 (8): 03120001. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001265.

Chandy, S. E., A. Rasekh, Z. A. Barker, and M. E. Shafiee. 2019. "Cyber-attack detection using deep generative models with variational inference." *J. Water Resour. Plann. Manage.* 145 (2): 04018093. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001007.

Douglas, H. C., R. Taormina, and S. Galelli. 2019. "Pressure-driven modeling of cyber-physical attacks on water distribution systems." *J. Water Resour. Plann. Manage.* 145 (3): 06019001. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001038.

Dzung, D., M. Naedele, T. Von Hoff, and M. Crevatin. 2005. "Security for industrial communication systems." *Proc. IEEE* 93 (6): 1152–1177. https://doi.org/10.1109/JPROC.2005.849714.

Erba, A., R. Taormina, S. Galelli, M. Pogliani, M. Carminati, S. Zanero, and N. O. Tippenhauer. 2020. "Constrained concealment attacks against reconstruction-based anomaly detectors in industrial control systems." In *Proc., Annual Computer Security Applications Conf., ACSAC '20*, 480–495. New York: Association for Computing Machinery.

Galloway, B., and G. P. Hancke. 2013. "Introduction to industrial control networks." *IEEE Commun. Surv. Tutorials* 15 (2): 860–880. https://doi.org/10.1109/SURV.2012.071812.00124.

Hassanzadeh, A., A. Rasekh, S. Galelli, M. Aghashahi, R. Taormina, A. Ostfeld, and M. K. Banks. 2020. "A review of cybersecurity incidents in the water sector." *J. Environ. Eng.* 146 (5): 03120003. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001686.

Humayed, A., J. Lin, F. Li, and B. Luo. 2017. "Cyber-physical systems security: A survey." *IEEE Internet Things J.* 4 (6): 1802–1831. https://doi.org/10.1109/JIOT.2017.2703172.

Kadosh, N., A. Frid, and M. Housh. 2020. "Detecting cyber-physical attacks in water distribution systems: One-class classifier approach." *J. Water Resour. Plann. Manage.* 146 (8): 04020060. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001259.

Kayan, H., M. Nunes, O. Rana, P. Burnap, and C. Perera. 2022. "Cybersecurity of industrial cyber-physical systems: A review." *ACM Comput. Surv.* 54 (11s): 1–35. https://doi.org/10.1145/3510410.

Kobayashi, T. H., A. B. Batista, A. M. Brito, and P. S. Motta Pires. 2007. "Using a packet manipulation tool for security analysis of industrial network protocols." In *Proc., 2007 IEEE Conf. on Emerging Technologies and Factory Automation (EFTA 2007)*, 744–747. New York: IEEE.

Krotofil, M., A. Cárdenas, J. Larsen, and D. Gollmann. 2014. "Vulnerabilities of cyber-physical systems to stale data-determining the optimal time to launch attacks." *Int. J. Crit. Infrastruct. Prot.* 7 (4): 213–232. https://doi.org/10.1016/j.ijcip.2014.10.003.

Labs, V. 2022. "Analysis of clop's attack on south Staffordshire water: UK." Accessed August 20, 2022. https://securityboulevard.com/2022/08/analysis-of-clops-attack-on-south-staffordshire-water-uk/.

Lantz, B., B. Heller, and N. McKeown. 2010. "A network in a laptop: Rapid prototyping for software-defined networks." In *Proc., 9th ACM SIGCOMM Workshop on Hot Topics in Networks, Hotnets-IX*. New York: Association for Computing Machinery.

Makropoulos, C., and D. Savić. 2019. "Urban hydroinformatics: Past, present and future." *Water* 11 (10): 1959. https://doi.org/10.3390/w11101959.

Murillo, A., R. Taormina, N. O. Tippenhauer, D. Salaorni, R. van Dijk, L. Jonker, S. Vos, M. Weyns, and S. Galelli. 2022. "High-fidelity cyber and physical simulation of water distribution systems. I: Models and Data." *J. Water Resour. Plann. Manage.* 149 (5): 04023009. https://doi.org/10.1061/JWRMD5.WRENG-5853.

Nikolopoulos, D., and C. Makropoulos. 2021. "Stress-testing water distribution networks for cyber-physical attacks on water quality." *Urban Water J.* 19 (3): 256–270. https://doi.org/10.1080/1573062X.2021.1995446.

Nikolopoulos, D., G. Moraitis, D. Bouziotas, A. Lykou, G. Karavokiros, and C. Makropoulos. 2020. "Cyber-physical stress-testing platform for water distribution networks." *J. Environ. Eng.* 146 (7): 04020061. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001722.

Ramotsoela, D. T., G. P. Hancke, and A. M. Abu-Mahfouz. 2019. "Attack detection in water distribution systems using machine learning." *Hum.-centric Comput. Inf. Sci.* 9 (1): 1–22. https://doi.org/10.1186/s13673-019-0175-8.

Rasekh, A., A. Hassanzadeh, S. Mulchandani, S. Modi, and M. K. Banks. 2016. "Smart water networks and cyber security." *J. Water Resour. Plann. Manage.* 142 (7): 01816004. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000646.

Rossman, L. A. 2000. *EPANET 2: Users manual*. Cincinnati: Water Supply and Water Resources Division, National Risk Management Research Laboratory.

Sánchez, H. S., D. Rotondo, T. Escobet, V. Puig, and J. Quevedo. 2019. "Bibliographical review on cyber attacks from a control oriented perspective." *Annu. Rev. Control* 48: 103–128. https://doi.org/10.1016/j.arcontrol.2019.08.002.

Sandaruwan, G. P. H., P. S. Ranaweera, and V. A. Oleshchuk. 2013. "PLC security and critical infrastructure protection." In *Proc., 2013 IEEE 8th Int. Conf. on Industrial and Information Systems*, 81–85. New York: IEEE.

Shapira, N., O. Ayalon, A. Ostfeld, Y. Farber, and M. Housh. 2021. "Cybersecurity in water sector: Stakeholders perspective." *J. Water Resour. Plann. Manage.* 147 (8): 05021008. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001400.

Shin, S., S. Lee, S. J. Burian, D. R. Judi, and T. McPherson. 2020. "Evaluating resilience of water distribution networks to operational failures from cyber-physical attacks." *J. Environ. Eng.* 146 (3): 04020003. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001665.

Tanenbaum, A. S., and D. J. Wetherall. 2010. *Computer networks*. 5th ed. Hoboken, NJ: Prentice Hall.

Taormina, R., et al. 2018. "Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks." *J. Water Resour. Plann. Manage.* 144 (8): 04018048. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000969.

Taormina, R., and S. Galelli. 2018. "Deep-learning approach to the detection and localization of cyber-physical attacks on water distribution systems." *J. Water Resour. Plann. Manage.* 144 (10): 04018065. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000983.

Taormina, R., S. Galelli, H. Douglas, N. O. Tippenhauer, E. Salomons, and A. Ostfeld. 2019. "A toolbox for assessing the impacts of cyber-physical attacks on water distribution systems." *Environ. Modell. Software* 112 (Feb): 46–51. https://doi.org/10.1016/j.envsoft.2018.11.008.

Taormina, R., S. Galelli, N. O. Tippenhauer, A. Ostfeld, and E. Salomons. 2016. "Assessing the effect of cyber-physical attacks on water distribution systems." In *Proc., World Environmental and Water Resources Congress 2016*, 436–442. Reston, VA: ASCE. https://doi.org/10.1061/9780784479865.046.

Taormina, R., S. Galelli, N. O. Tippenhauer, E. Salomons, and A. Ostfeld. 2017. "Characterizing cyber-physical attacks on water distribution systems." *J. Water Resour. Plann. Manage.* 143 (5): 04017009. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000749.

Tsiami, L., and C. Makropoulos. 2021. "Cyber-physical attack detection in water distribution systems with temporal graph convolutional neural networks." *Water* 13 (9): 1247. https://doi.org/10.3390/w13091247.

Tuptuk, N., P. Hazell, J. Watson, and S. Hailes. 2021. "A systematic review of the state of cyber-security in water systems." *Water* 13 (1): 81. https://doi.org/10.3390/w13010081.

Urbina, D. I., J. A. Giraldo, N. O. Tippenhauer, and A. A. Cárdenas. 2016. "Attacking fieldbus communications in ICS: Applications to the SWaT Testbed." In *Proc., Singapore Cyber-Security Conf. (SG-CRC) 2016*, 75–89. London: IOS Press. https://doi.org/10.3233/978-1-61499-617-0-75.