Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

Curse of System Complexity and Virtue of Operational Invariants: Machine Learning based System Modeling and Attack Detection in CPS

Muhammad Omer Shahid
CUST
Islamabad, Pakistan
omer.tcp@gmail.com

Chuadhry Mujeeb Ahmed
University of Strathclyde
Glasgow, United Kingdom
mujeeb.ahmed@strath.ac.uk

Venkata Reddy Palleti
IIPE
Andhra Pradesh, India
venkat_palleti.che@iipe.ac.in

Jianying Zhou
SUTD
Singapore
jianying_zhou@sutd.edu.sg

***Abstract*—Cyber Physical Systems (CPS) security has gained a lot of interest in recent years. Different approaches have been proposed to tackle the security challenges. Intrusion detection has been of most interest so far, involving design-based and data-based approaches. Design-based approaches require domain expertise and are not scalable, on the other hand, data-based approaches suffer from the lack of real-world datasets available for specific critical physical processes. In this work, a data collection effort is made on a realistic Water Distribution (WADI) test-bed. Collected data consists of both the normal operation as well as a range of attack scenarios. Next, machine learning-based system-modeling techniques are considered using the data from WADI. It is shown that the accuracy of system model-based intrusion detectors depends on the model accuracy and for non-linear processes, it is non-trivial to obtain accurate system models. Moreover, an operational invariants-based attack detection technique is proposed using the system design parameters. It is shown that using a simple rule-based anomaly detector performs better than the complex black-box data-based techniques.**

***Index Terms*—CPS Dataset, Machine Learning, Attack Detection, ICS Security, CPS Security**

I. Introduction

The core composites of a Cyber Physical System (CPS) are one or more physical processes controlled using computing systems [1] referred to as Programmable Logic Controllers (PLCs) and Remote Terminal Units (RTUs). Critical infrastructures make use of advances in communication technologies and realize the remote control and operations enabled through the use of PLCs and RTUs. The focus in this article is on Critical Infrastructure (CI), specifically water distribution systems. CI heavily relies on automation to enable effective control and monitoring of the physical processes. However, the same networks simultaneously expose the system to malicious actors. Securing CPS is challenging and different from the pure IT systems in different ways [2], [3]. Recently there is a lot of attention being paid to develop defense technologies for CI. To contain and react to an attack it is highly important to detect it first. Therefore, researchers are designing anomaly detection systems either based on the design of the process or the data from the Physical process [4]–[6].

Recent literature on defensive approaches often look at CPS from control-theoretic perspective [7]–[12]. Such approaches depend on the principles of closed-loop feedback control to understand how the system deviates under attacks from normal behavior. These studies provide insights and mathematical bounds to the attacks but are limited due to theoretical limitations of underlying models of the physical process. It turns out that it's a non-trivial problem to find accurate system models for a complex real-world system. One question we are addressing in this work is, *How precise a system model we can get using standard algorithms?* In contrast, experimental and design-based approaches [13], [14] use testbeds to demonstrate the effectiveness of attack detection methods. Such experimental studies point to the importance of testbeds for CPS security research.

System model-based techniques have been common in the literature but those techniques have limitations. There have been extensive studies on the limitations of model-based attack detectors [4], [15], [16]. Alternatively, other research efforts focus on the use of machine learning to generate system models and also design anomaly detectors. In this article, we highlight the limitations of machine learning-based techniques. The use of machine learning to create anomaly detectors becomes attractive with the increasing availability of data and advanced computational resources. However, the data-based techniques rely on a rich dataset representing a real-world scenario. Such datasets are not easily accessible to academia. Our goal is to create a unique set of data that is 1) accessible and 2) represents real-world settings. In this article based on our experiments in a water distribution testbed, we have collected data for different states of the physical process under normal and attack conditions.

There are some other efforts for data collection beyond iTrust labs at Singapore University of Technology and Design, but those still have few limitations. An interesting effort in electric grid testbed is simulation-based Softgrid testbed [17] but there is no data generation and sharing. For an ICS testbed in CPS [18] authors highlighted that their prototype lacked the collection and distribution of data as it involves a manual process requiring time and resources. [19] presented simulated IEC61850 traffic and no information regarding the real process and dynamics. Previous research studies have tried to collect

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

data from CPS settings but lack some desired features. We highlight a few of those,

- Simulated Data: Most of the datasets available are generated using simulated models [19]. The lack of real-world scenarios prompted us to do this work.
- Only Network Traffic Data: Previous data collection efforts only focused on the network traffic and as mentioned those were too simulated most of the time. One of the recent studies collected the network traffic from a realistic electric traction substation [20]. There is a lack of process data available.

Contributions: Our contributions are multi-fold. We run extensive experiments and attack scenarios on a real water distribution network and collect data to share with academia and industry. We have focused on the process data from the sensors and actuators and tried to run the normal process for a range of different configurations. On the other hand, we have executed different attacks for several days and have collected and shared the data publicly [21]. Next, we used machine learning techniques to come up with system models and point on the limitations of such techniques, especially for complex CPS networks. We proposed an operational invariant technique in contrast to the process invariants (relies on the system model) and performed an extensive analysis on 15 different attack scenarios and report the detection results.

II. Background: WADI Testbed and Data Collection

A. Water Distribution Networks

Water distribution networks are an important part of any city and are often spread over several kilometers. These systems are vulnerable to different kinds of attacks [22]. Programmable Logic Controllers (PLCs) and Remote Terminal Units (RTUs) are used for efficient monitoring and control of these systems. In order to operate these systems SCADA systems are used for centralized supervision. PLCs collect data from the sensors, this information is used to compute control actions to be sent to the actuators. This automation relies on the cyberinfrastructure to exchange information between the controllers and SCADA. The addition of this cyber layer makes the water distribution network vulnerable to cyber-physical attacks [23]. According to a report by ICS-CERT, several attacks have occurred against water utilities in USA [24]. The critical nature of water distribution infrastructures makes them an attractive target for hackers and terrorist activities. Therefore, it is important to ensure both the physical and cyber security of these systems. Therefore, it is very important to detect such types of attacks so as to avoid any economic loss or service disruptions. The goal of this paper is to design a detection mechanism using a realistic data set obtained from an operational testbed. The following sections describe the architecture of WAter DIstribution (WADI) system testbed and the data collection.

The WADI testbed is a fully operational physical testbed that represents a scaled-down version of a real urban water distribution ICS with the capacity to provide a distribution output of 10 gallons/minute [25]. As shown in Figure 1, the testbed is comprised of three stages: a primary grid (P1), a secondary grid (P2) containing an elevated reservoir (ER); and a return water grid (P3). P1 is comprised of two 2500 litre capacity raw water tanks that are fed by two sources; a raw water inlet valve and return water grid P3. Chemical dosing pumps are installed to maintain a consistent water quality input to these tanks and a water level sensor is installed in each tank. Water quality is monitored by sensors in all three stages with measurements being made on water conductivity, turbidity, PH and Oxidation Reduction Potential (ORP). In addition, two contaminant sampling stations, P2A and P2B, are installed in the testbed to measure water quality parameters before its delivery to the consumer tanks. The P2 stage consists of two elevated tanks and six consumer tanks and it should be noted that the dynamics of the whole system is driven by the preset demand settings of the consumer tanks. Based on these presets, water flows from the elevated tanks into the consumer tanks at a certain rate and once the consumer tanks are filled, water drains into the return grid (P3), which in turn supplies the primary grid (P1).

B. Data Collection

WADI is a water distribution testbed, therefore, treated water is distributed to the consumer tanks and the dynamics of the system depends on the demand pattern of these tanks. The demand pattern is generated for each consumer tank for each hour over 24 hours of a day. The demand profile of each consumer tank consists of low to high peak demand scenarios. The testbed ran for a total of 16 days in which 14 days consists of normal operation data and the remaining two days are attacked data. The data is collected every second for 16 days.

Initial 14 days the system is operated under normal conditions i.e, without any attacks. During the next two days, 15 different types of attacks are launched on the testbed bases on the attacker's intention. In these experiments. an insider attacker profile is assumed in which he/she has the process, communication knowledge, and access to the communication networks. Attacks are designed based on the attacker's intention. For example, an attacker intends to cut-off water supply to consumers in a water distribution system and overflow water storage tanks, etc. To realise the attacker's intentions, he/she launches malicious attacks on sensor measurements and actuators. Further, while launching attacker also tries to maintain the process safety requirements intact so that he can move the process into an abnormal state. The attack targets and detailed attack description are shown in Appendix A and the data is publicly available from iTrust website at Singapore University of technology and Design [1].

III. System Modelling

A system model of a physical process helps to model the process dynamics. Few approaches are based on the control

[1] https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

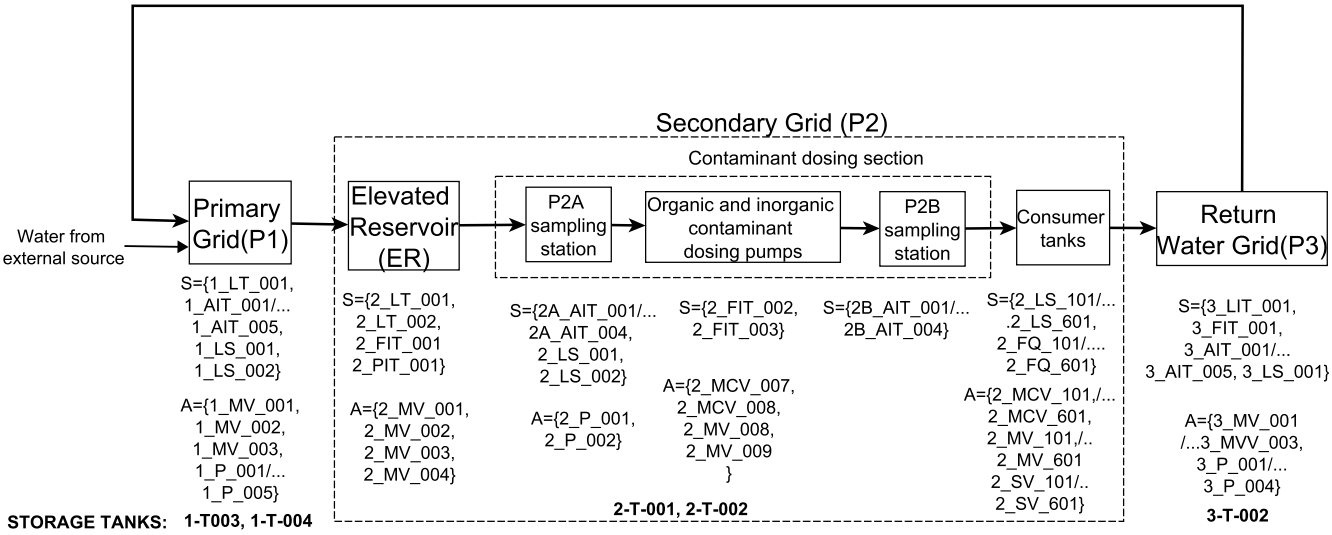

Fig. 1: Three stages of Water Distribution System

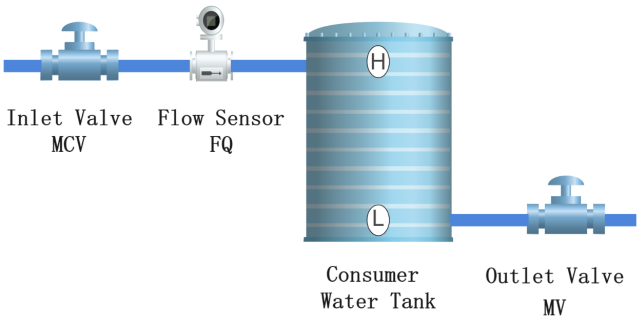

Fig. 2: Stage 2 of WADI: Example of a Consumer Tank Process.

systems theory [4], [16] and more recent works use the machine learning-based approaches [3]. In this work, we are going to rely on data-based machine learning techniques to generate the system models from the data and then validate those. Recently, system model-based approaches have been used for attack detection as well, however, we will show that those techniques are limited by the accuracy of the model. In real-world non-linear systems, it becomes challenging to obtain accurate models and hence difficult in using those for attack detection.

A. System Model based Attack Detection

Figure 2 shows a water tank and the inflow and outflow of the water is being controlled by the motorized valve (MV) at the input and at the output respectively. The idea is to model this inflow and outflow by considering the physical principles and the design of the physical process. For a tank, we know that the rate of change of water inside the tank is equal to the difference between water flowing into the tank and water flowing out from the tank with respect to time. We can represent this using the mass-balance equation such as,

$$\frac{dV}{dt} = Q_{in} - Q_{out}$$

$$\frac{dh}{dt} = \frac{Q_{in} - Q_{out}}{A} \quad \text{since} \quad V = A \times h, \tag{1}$$

where V represents the volume of the tank, A is the cross-sectional area of the tank, and h is the height of the water inside the tank, (1) provides a linear equation, we can see the term $[Q_{in} - Q_{out}]$ represents the water flow which depends upon the PLC control actions implemented via inlet MCV and outlet MV. From Figure 2, it can be seen that using the height and diameter of the tank from design documents, it is possible to figure out the volume and the cross-sectional area of the tank. Let us consider the state of the physical process as the height of the water inside the tank. Then the solution of this equation gives us the following result.

$$x_{k+1} = x_k + u_k,$$

where u_k is the PLC control action. Here x_k represents the water level in the tank at time k. The control action u_k can be either open/close (for the motorized valve) or on/off (for the pump). Similarly, we can describe the sensor state and we can get the set of system equations. However, it is difficult to obtain and scale the system model from the design as it requires design specifications and domain expertise, therefore, in this work we have considered data-based approaches using machine learning techniques.

B. Attack Detection Framework

A general model-based attack detection framework has two major components, 1) system model and estimation and 2) a threshold-based detector.

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

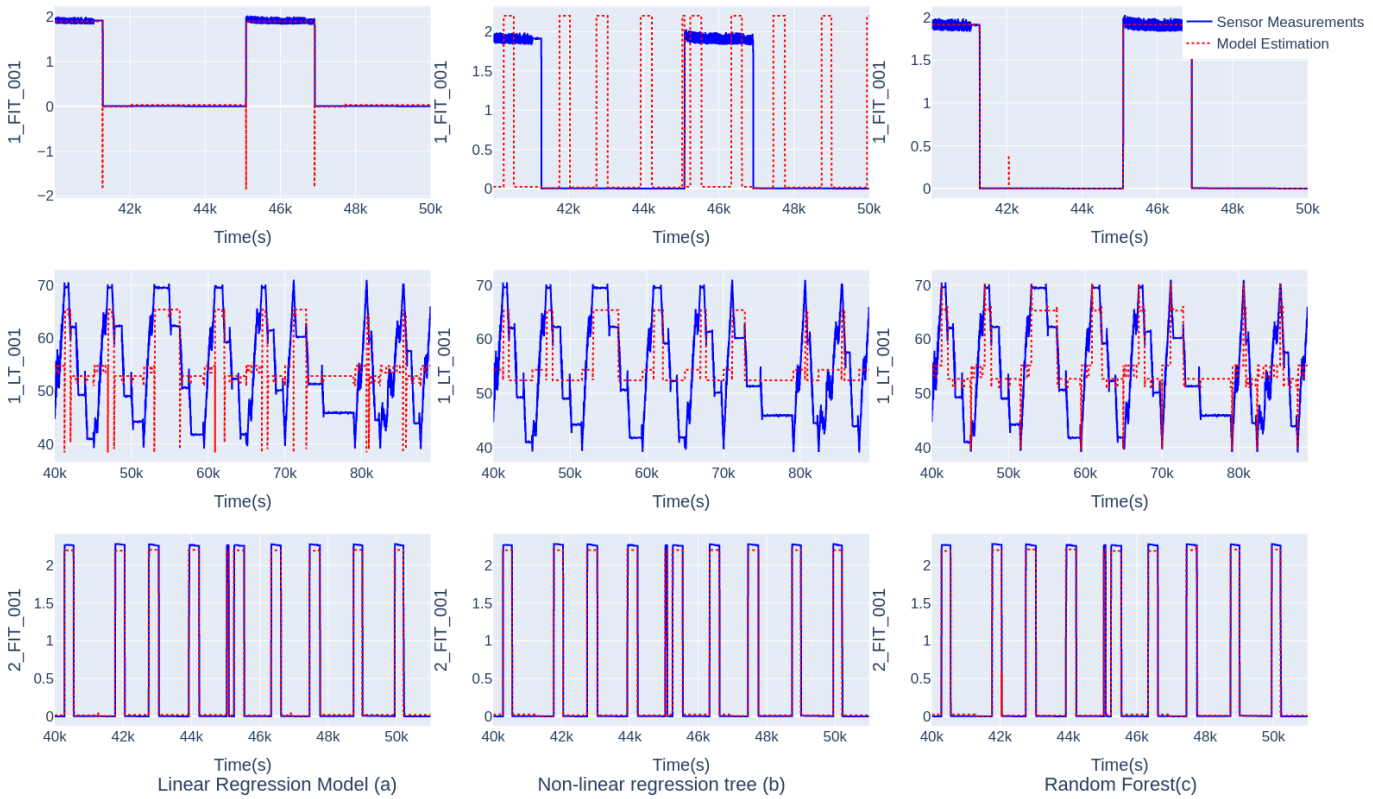

Fig. 3: Linear and non-linear modeling techniques used to obtain the system models for three sensors in Stage 1 of the WADI testbed.

System Model and Estimation: The idea of obtaining a system model is explained in the previous section. The system models can be obtained either using data based techniques or the first principles [4], [26]–[28]. Using the system model it is possible to estimate the states of the system and ultimately estimate output from a sensor ($\hat{y}_k$). A residual vector is calculated by taking the difference between the sensor measurements and estimated sensor output as,

$$r_k = y_k - \hat{y}_k. \tag{2}$$

Where r_k is the residual vector.

Threshold based Detector: To detect the presence of an attack, the residual vector is tested against a predefined threshold designed for a particular false alarm rate. We can create a threshold for the residual distribution and if the values of residual are outside that threshold declare it under an attack,

$$|r_k| > \tau, Alarm = True. \tag{3}$$

Where τ is a threshold and $|r_k|$ is the absolute value of the residual. There have been studies on optimizing the parameters of different stateful and stateless detectors [4], [29]. A wide variety of algorithms exist to chose the best threshold value to maximize the attack detection rate and minimize the false alarm rate. However, the success of these system model techniques is limited by the accuracy of a model. For some physical process state variables, it is possible to get accurate models but for some, it is not, as we will see in more detail in the following sections.

C. Different Modeling Techniques

As stated earlier, in this work we explore the application of machine learning-based techniques to obtain the system model using the data collected from the WADI testbed under the normal operation. Moreover, we compared different machine learning approaches to assess the model accuracy. A separate model for each stage of the testbed is obtained. Let's consider stage 1 of the WADI testbed. The sensors and actuators considered in this stage are level sensor (1_LT_001) on the water tank, flow sensor at the inlet of the tank (1_FIT_001), flow sensor at the outlet of the tank (2_FIT_001) and motorized valves at the input of the tank (1_MV_001, 1_MV_004), pumps at the outlet of the tanks (1_P_005, 1_P_006), respectively. Note, that there are two pumps on the output of the tank in stage 1 because one of the pumps is the backup in case the other gets faulty.

The first approach was to apply Linear Regression Model (LRM) due to its being simple and explainable.

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4. \tag{4}$$

Where y is the dependent variable and x_i are the explanatory variables. b_i are the coefficients of x and a is the intercept.

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

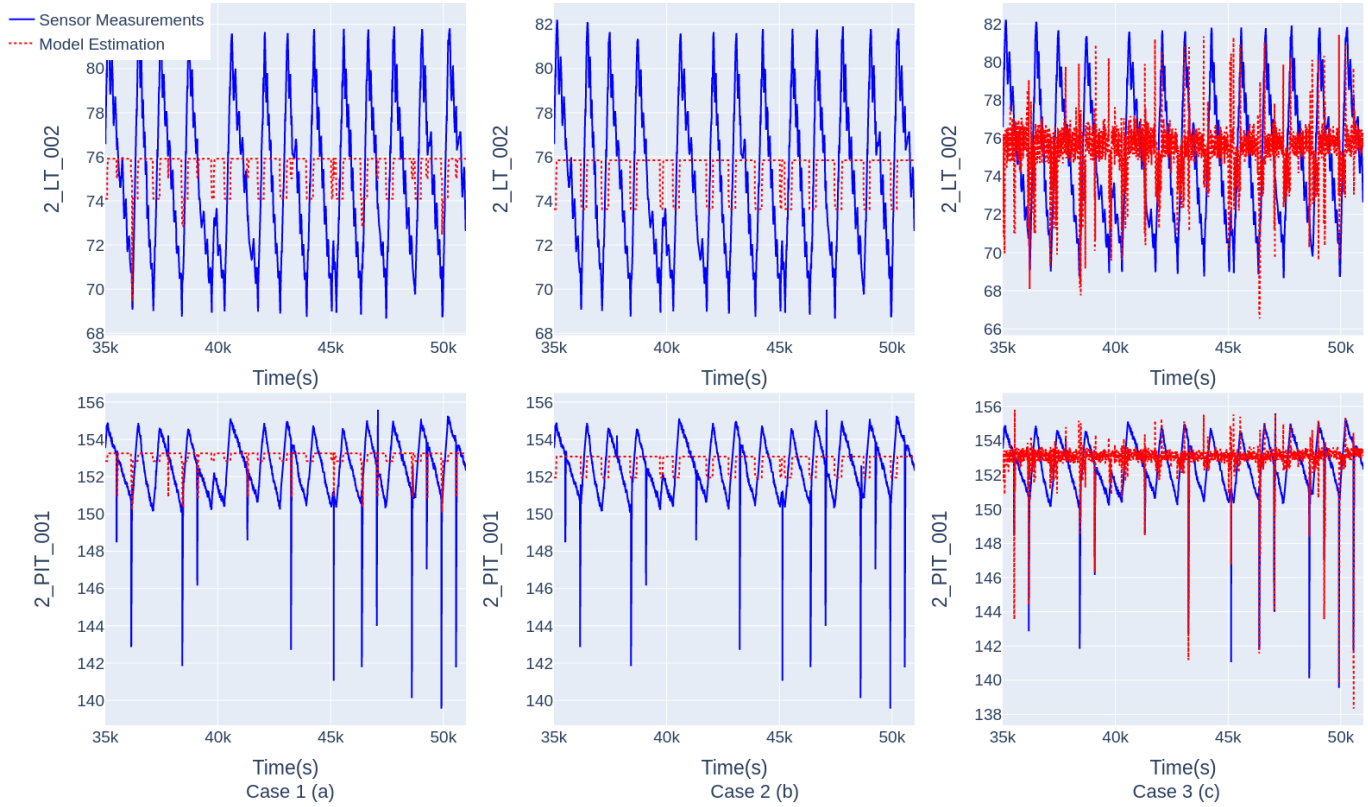

Fig. 4: Relationship of different input vectors on the modeled sensor output. For the two sensors in stage 2 it is seen that for three different input vectors, it is difficult to get precise models using the random forest regression technique.

In our model, we take data from a sensor as y and four actuators as x_i, to obtain the coefficients and intercept value. For the three sensors in the stage 1 LRM coefficients and intercept as defined by Eq. (4) are provided. x_1, x_2, x_3, x_4 represents 1_MV_001, 1_MV_004, 1_P_005, 1_P_006, respectively. Note that in the following for all coefficients b_4 is 0, the reason for that is 1_P_006 is the backup pump and it never gets to turn ON during the data collection process as it turns ON only if 1_P_005 is faulty.

Sensor: 1_FIT_001
Intercept: -1.7986565219245392
Coefficients: [1.86663684, -0.03652458, -0.00212879, 0.]

Sensor: 1_LT_001
Intercept: 39.66108207833245
Coefficients: [1.89978019, 12.55035734, -1.25784017, 0.0]

Sensor: 2_FIT_001
Intercept: -2.1478883186779028
Coefficients: [-0.00657396, -0.01372888, 2.18690766, 0.0]

Next, we validate the obtained linear model by using actuator data and estimating the sensor measurements. The performance of the model is assessed using the residual, i.e., the difference between the sensor measurement and sensor estimates $r_i = y_i - \hat{y}_i$ and mean square error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

Figure 3 (a) shows the results of the LRM obtained for three different sensors at the stage 1. We can see that the models for two flow sensors are accurate as the sensor estimates are accurate, while for the level sensor, estimates deviate from the actual sensor measurements. Table I shows inputs used to obtain the model and sensor outputs with their respective MSE. From Table I it can be seen that LRM results in a high error for the level sensor, in agreement with the pictorial view of the Figure 3(a). Motivated by the poor result using the linear model, we used non-linear regression trees and random forest algorithms to capture any non-linearity in the data. Figure 3(b,c) shows the results, where it can be observed that there are no gains made in terms of model accuracy. Model accuracy would have improved using non-linear models if the data was not linearly separable but it turns out that the data is not separable linear or otherwise [30]. However, it can be seen in Figure 3(c) that the estimator is better able to trace the pattern in the level sensor measurements but MSE is still high to render the model useless. Upon further investigations, it is highlighted that although flow processes are well defined and

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

Inputs	Outputs	LRM	NLRT	RF
1_MV_001 , 1_MV_004 , 1_P_005 , 1_P_006	1_FIT_001	0.0175	0.0049	0.0016
	1_LT_001	44.0786	42.9155	42.3352
	2_FIT_001	0.04009	0.04008	0.03742

TABLE I: Mean squared error of output sensor of stage 1 models using normal data.

Case No	Inputs	Outputs	MSE : 2_LT_002 , 2_PIT_001
01	2_MCV_007 , 2_MV_003 , 2_MV_004 , 2_MV_006 , 2_MV_009	2_LT_002 , 2_PIT_001	10.8193 , 2.7808
02	2_MV_003 , 2_MV_004		10.8779 , 3.36599
03	2_MV_003 , 2_MV_004 , 2_MV_006 , 2_P_003		12.9912 , 2.62007

TABLE II: Effects of inputs on model accuracy. Three different use cases showing combinations of inputs on the model accuracy.

Senor	T1	T2	T3	T4	T5	T6
2_FQ_x01	0.018	0.014	0.013	0.015	0.017	0.015

TABLE III: Mean squared error of output sensor of stage 2 models normal data (consumer tanks where x= 1,2,3,4,5,6).

could be modeled well even with the simple linear models, the level sensor data is irregular due to the physical irregularities in the testbed and it is hard to obtain an accurate model, which is in line with the earlier works using the control theory system identification on the same data [5].

From Table I it can be seen that among all the models Random Forest (RF) performs the best. Therefore, it will be used to model the other stages in the testbed. Table III shows the MSE of the models for the flow sensors installed in six consumer tanks using the RF algorithm. Consumer tank models are quite accurate and will be used for attack detection as discussed in Section V.

D. Complexity: Effect of Inputs on Model Accuracy

Since we have concluded that the Random Forest is the best to create system models, therefore, for stage 2 of the WADI testbed, we will use Random Forest(RF) with n_estimators (number of trees) size = 100 to create different system models. Having chosen a regression algorithm, here we would like to test the effect of choosing input combinations on the model accuracy for the sensor outputs. WADI represents a realistic water network and has a complex structure with multiple inputs and outputs. Stage two of the WADI has a combination of booster pumps and gravity-driven flow due to elevated tanks based on the demand pattern of the consumers. To assess the effects of input combinations on the model, we have created models with eleven different combinations of inputs and outputs. From which we have selected three cases to explain the idea as shown in Table II. This analysis is done for the two most irregular outputs, i.e., the level sensor (2_LT_002) and the pressure sensor (2_PIT_001) as shown in Figure 4. It is concluded that using more inputs does not behave any better than using simply two inputs at the input and output of the elevated tank as is the case 2. The models are so bad that those are simply useless for threshold-based attack detection. But we have only shown the worst-case scenarios here, there are other sensors for those we have obtained accurate models as shown in Figure 3. In the following, we propose another technique based on the operational invariants instead of the model-based attack detection.

IV. Invariants based Attack Detection

Definition: An “invariant" is a condition among physical properties of a process that must hold in ICS is reflected to be in that state by the data [13].

State Space based IDS: Based on the design of the physical process a state-space could be learned. The idea is to extract internal relations among system variables and derive a graph model to describe valid system states [31].

Process Invariants based IDS: Consider an example of a water tank. A process invariant is derived from the dynamics of the water flow process into and out of the tank [13]. Several design parameters were defined from the system dynamics.

Control Invariants based IDS: This type of IDS is based on deriving and monitoring Control Invariants. It is a method to extract such invariants by jointly modeling physical properties, control algorithms, and the laws of physics. These invariants are represented in a state-space form, which can then be implemented and inserted into the control program binary for runtime invariant check [14].

These process invariant-based techniques as explained above depends on the model of the physical process and exploring the state space of the whole process. However, one advantage of those methods is that they can discover the state space of the process and might result in a better IDS but the key challenge is that the physical process dynamics tend to be complex, therefore, it is really hard to obtain a precise system model.

The above methods need to track the whole state space in real-time that can be a difficult task.

Operational Invaraints Operational invariants can be derived from the design of the process. There does not need to exhaustively look at all the state space of the physical process. For example, for a water tank system the operational invariants would be the minimum and maximum limits of the tank for that it can hold the water, let us call those as Low and High values, respectively.

A. Motivating Example

Consider the example of a consumer in the context of the WADI testbed as shown in Figure 2. On the tank itself, there

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

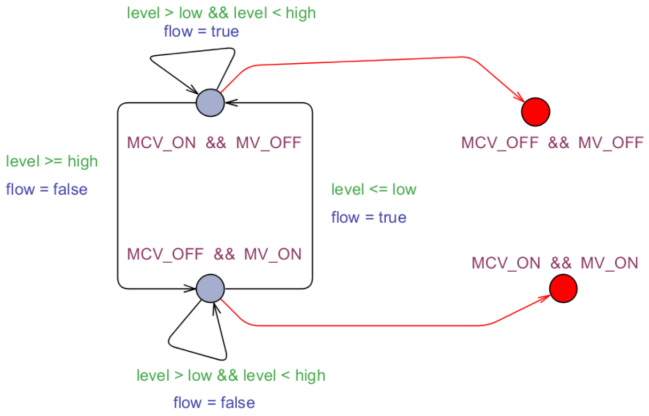

Fig. 5: State machine representation of the process states in a consumer tank.

are two limits high (H) beyond which the tank will flood and low (L) beyond which we do not have any water left for use. Outlet valve MV represents the consumer's usage, while inlet valve MCV is operated based on the demand pattern of the consumer. Given that the user's demand pattern could be complex and at best a stochastic process, it could be challenging to exhaustively obtain the state-space model. Whilst operational invariants provide us a simple set of rules to figure out if something is wrong. For example, if the tank is filled with water up to the 'H' mark then any more water flowing in would result in overflow/flooding. This could be the attacker's intention as we will see in the attack table in the following section. Based on the operational limits we can come up with a rule that *if the level of water is equal to H then MCV should be turned OFF*.

Figure 5 is a Finite State Machine (FSM) for consumer tanks from stage 02 of WADI. there are six consumer tanks. Each has one inlet valve(MCV), one flow sensor(FQ), one water tank, and one outlet valve (MV). This FSM shows the relationship between the inlet valve outlet valve flow sensor of a consumer tank. The possible operational invariants under normal scenario are expressed in equations IV-A and IV-A.

$$(MCV\_ON \wedge MV\_OFF) \wedge (Level > L \wedge level < H) \implies Flow = True \quad (6)$$

$$(MCV\_OFF \wedge MV\_ON) \wedge (Level > L \wedge level < H) \implies Flow = False \quad (7)$$

Further, a lot of real-world attacks can be detected using such operational invariant monitors. Take the example of a recent attack on the water treatment system of Oldsmar, Florida in February 2021. The attacker took over the SCADA system and turned ON the chemical dosing pump to increase the level of Sodium Hydroxide (NaOH) 100 times its normal levels [32]. Considering the use of operational invariants we could have interlocked the chemical dosing pump ON/OFF with the operational ranges of NaOH, for example, the chemical dosing pump can only be ON if NaOH concentration is below the low (L) level in the water. A typical example of a pH dosing system is shown in the code listing below. For an operational invariant-based technique we can interlock the dosing pump ON/OFF actuation with the measurements from a pH sensor.

```
# HCl is used to comply with regulations
if pH_value >= 7.05:
    dosing_pump = ON
elif pH_value <= 6.95:
    dosing_pump = OFF

```

Listing 1: pH dosing operational parameters for a typical water treatment system.

V. Results and Discussion

In this section, the attack detection mechanisms are discussed for each attack launched on the WADI testbed. As described in the previous sections the model estimation techniques and invariants-based approach are used to detect different types of attacks. As an illustration, the detection mechanism of attack 5 of table IV in A is detected using model estimation. Later, we will describe all the attacks and their detection mechanisms.

Figure 6 shows the plot of results of model created for consumer tank 1. The left-hand column shows the consumer tank under normal operation. The top-left hand plot shows the normal flow sensor measurements along with the estimate from the obtained system model and the bottom plot shows the residual. On the right-hand column, attack-5 is shown with the circle on the attack duration which starts from the index 63041 to 63891. It can be seen that the residual under attack deviates from the normal pattern and could lead to the detection of the attack. Further, this attack can also be detected using operational invariants which will be explained in the next subsection.

Attack Detection Performance:

- *Attack 1:* In this attack, the attacker maliciously turned on 1_MV_001 in stage 1 of the process in order to realize the overflow of the raw water tank. This attack can be detected using both operational constraints and model estimation. Attack 01-subplot in Figure 7 shows the level sensor 1_LT_001 reading above 70% which is High setpoint. The operator can easily detect the attack once the level exceeds the High setpoint. Further, Figure 8 shows the residual error between the sensor measurement and model estimation and one can observe that these residuals deviate significantly from the normal thresholds.
- *Attack 2:* The attacker's intention is to increase the chemical dosing levels in the primary grid. The attacker targets the flow meter 1_FIT_001 and sends a false reading to the controller so that the chemical dosing pump starts injecting the chemical into the water. This attack can be detected using the following invariant between the inlet valve 1_MV_001 and 1_FIT_001 as shown in

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

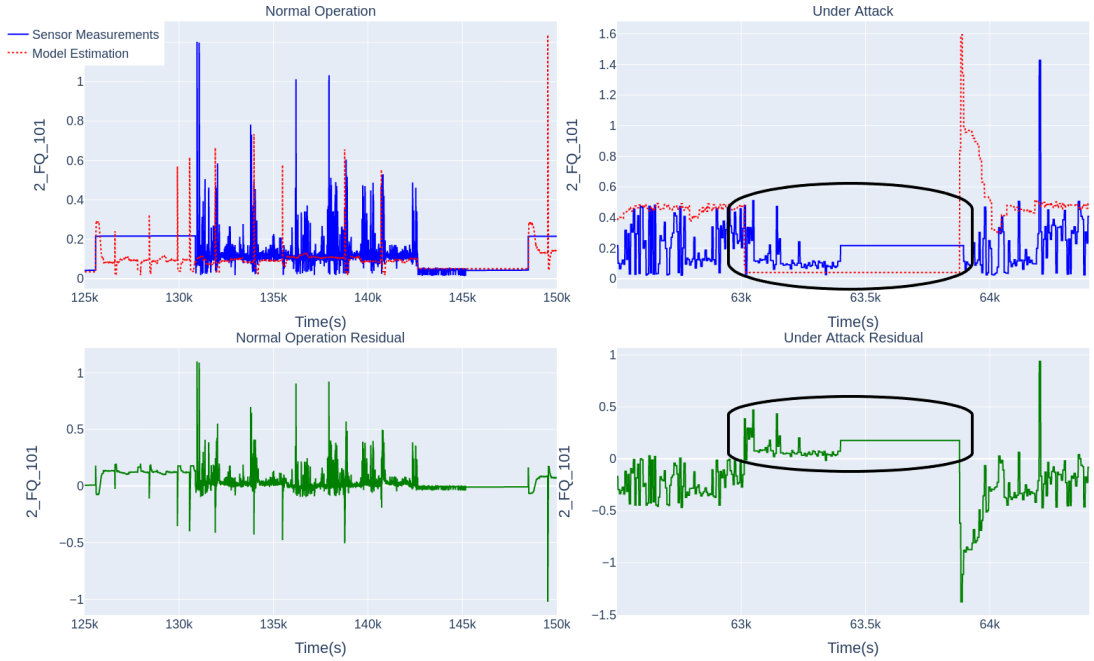

Fig. 6: Consumer tank 1 under normal operation and attack. This attack corresponds to attack 5 of table IV. The black circled area is the duration of the attack

.

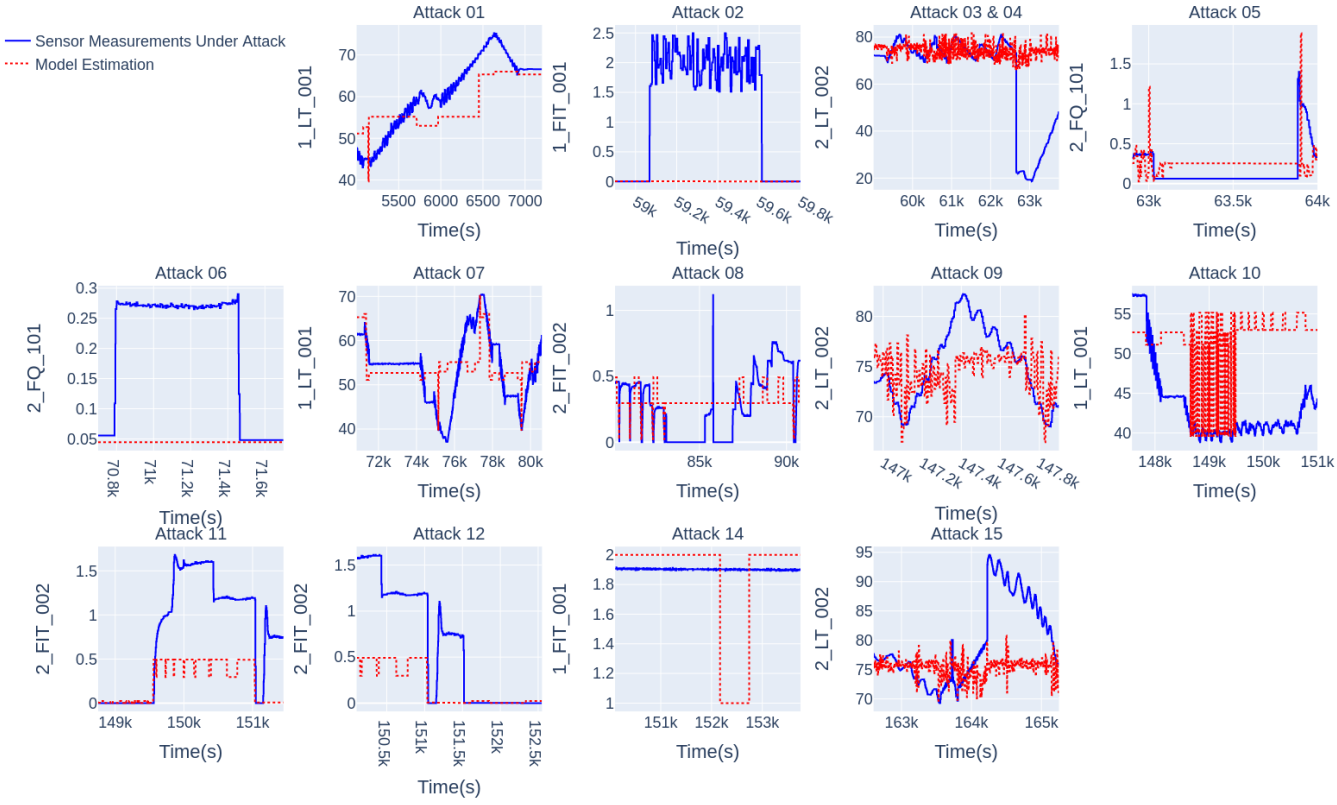

Fig. 7: Visual inspection of attacks on different stages of WADI. Sensor measurements (attacked) and model estimates for those sensors are shown.

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

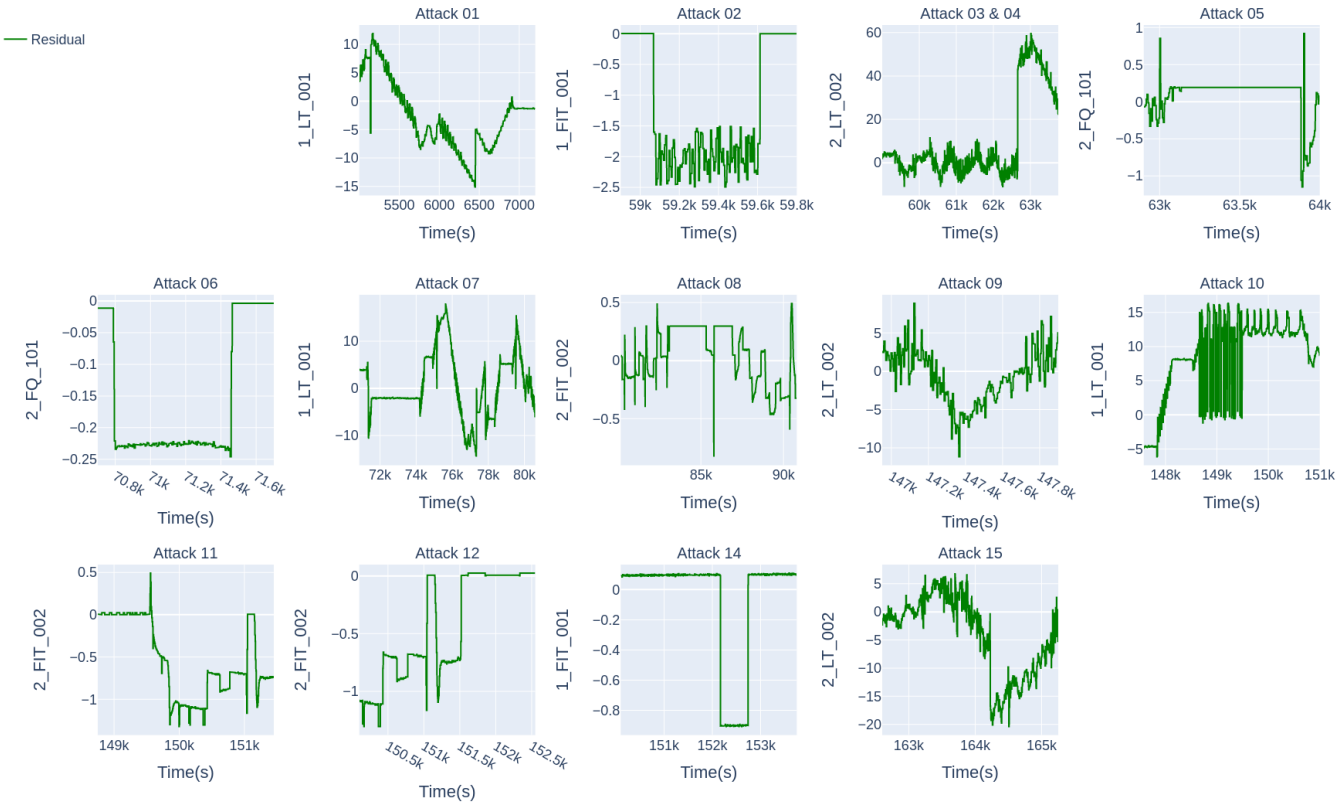

Fig. 8: Residual signal for all the sensors under attack as shown in Figure 7.

equation 8. It can also be observed from Figure 7 that the model estimation is showing the zero reading in 1_FIT_001. This is due to the reason that the inlet valve 1_MV_001 is closed.

$$1\_MV\_001 - OFF \implies 1\_FIT\_001 = 0$$
$$1\_MV\_001 - ON \implies 1\_FIT\_001 > 0 \quad (8)$$

- *Attack 3&4:* In this attack, the attacker's intention is to drain the elevated reservoir tank. In order to achieve that a stealthy attack is performed on sensor 2_LT_002 as shown in Figure 7. This attack can not be detected during the attack period, however, once the attack is stopped it can be detected using the tank level set points. As it can be seen from Figure 7 the tank level dropped to 20% which is the Low Low set point. If the stealthy attack continues to operate for some more time, the tank level goes empty and may cause damage to the outlet pump. A similar observation can be made from Figure8 that the error deviated from normal operation after stopping the attack. However, the attack is not detected during the attack period as the error is within the normal threshold.
- *Attack 5:* This attack can be detected using model estimation and finite state machine. The model estimation detection technique is already explained above. Here, a finite state machine is used to detect the attack. This figure explains the logical relationship between the inlet valve (MCV), outlet valve (MV) and flow sensor. The logical expression in Equation IV-A explains that the flow sensor (2_FQ_101) will be true only when MCV is open and MV is in the closed position and level is in between L and H. Hence the attack is detected.
- Attack 6: The similar arguments hold for this attack as explained in attack 5. However, This attack can be detected using the expression as shown in Equation IV-A.
- Attack 7: In this attack, the attacker tried to manipulate the water quality sensor (1_AIT_002). During the attack, the raw water tank starts draining because of bad water quality. Therefore the attack can be detected using model estimation as shown in Figure 7. Further, in Figure 8 it is observed that the error on 1_LT_001 deviated from normal.
- Attack 8: In this attack, the attacker wants to steal water from the main pipeline section by maliciously opening the valve 2_MV_007 to 30% position. The impact of the attack is shown on the sensors 2_FIT_002 and 2_PIT_002. The model-based detection mechanism can detect this attack using 2_FIT_002 sensor readings. In Figure 7 and a subplot of attack 8, model estimation shows a significant deviation from the normal sensor measurements.
- Attack 9: According to the design of the system only one pump (1_P_005) should be operated for pumping water from the raw water tank to the elevated reservoir tank. However, the attacker intentionally turns on the stand-by pump (1_P_006) to increase the level in the elevated tank

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

or damage the pipeline. As can it can be seen from figure 7 that the level shoots up to 80% of the tank which is the High High setpoint. This attack is detected through model estimation as the error in Figure 8 deviates from the normal values.

- Attack 10: The attacker intends to damage the valve (1_MV_101) and pump (1_P_005) by repeatedly turning ON and OFF. Further, the attacker wants to drain the elevated reservoir tank located downstream. This attack can be detected through an operational invariant. According to the process design, 1_MV_001 is ON whenever the level of the tank falls below 40% of the tank height. Further, the valve remains open until the level of the tank reaches 70%. The following invariant indicates that the valve is ON when the level of the tank is between 40% and 70%. Therefore the attack violates this invariant and hence detected.

$$1\_MV\_001\ ON \implies 40 \leq 1\_LT\_001 \leq 70 \quad (9)$$

- Attack 11: In this attack, the attacker tries to steal water from the main pipeline by opening the bypass valve 2_MV_007 from 0% to 50%. This attack is detected by sensor 2_FIT_002 as shown in the subplot of attack 11 of Figure 7. The model estimation measurement shows the significant deviation from the actual sensor measurements of 2_FIT_002.
- Attack 12: This attack is much similar to attack 11, however, the attacker produces water leakage by gradually opening the 2_MV_007 from 0 to 100% with an increment of 10%. This attack can also be detected by sensor 2_FIT_002 as shown in figure 7.
- Attack 13: Attacker intends to reduce the booster set point pressure, this causes intermittent water supply to consumers. We fail to detect this attack using either model-based detector because none of the associated sensors picked up this attack. Moreover, operational invariants could not be applied as this behavior could well be in normal ranges. This can be considered as a form of a stealthy attack.
- Attack 14: In this attack, the attacker maliciously turned OFF the chemical dosing pumps, namely, 1_P_001 and 1_P_003 in order to stop the chemical dosing to the raw water tank. The model-based estimation is used to detect the attack as shown in figure IV on the sensor 2_FIT_001.
- Attack 15: In this attack, the attacker's intention is to overflow the elevated reservoir tank (2_LT_002) by launching a stealthy attack. This attack could not be detected during the attack period as it is evident from the Figures 7 and 8. Once the attack is stopped the actual level of the tank is reached to 95% as shown in Figure 7. This shows that the attacker's intention is achieved. However, deviations are observed for quite some time between model estimation and sensor measurement even after the end of the attack.

VI. Conclusions

WADI testbed is used as a case study to collect data under normal and attack scenarios to be used for cyber security studies of CPS. Data-based techniques being popular nowadays, are tested for the purpose of modeling the physical systems. While those techniques perform reasonably well for some systems but have limitations to model accurately the complex processes. Moreover, it is shown that using the simple rules derived from the operational constraints of the system can be useful for attack detection, as was the case in the recent breach of the water system in Florida [32]. A detailed discussion is presented on attack detection using a combined system model and operational invariants-based approach.

References

[1] E. A. Lee, Cyber physical systems: Design challenges, in: EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2008-8, http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS- 2008-8.html, 2008.

[2] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, S. Sastry, Challenges for securing cyber physical systems, in: Workshop on future directions in cyber-physical systems security, 2009, p. 5.

[3] C. M. Ahmed, G. R. M R, A. P. Mathur, Challenges in machine learning based approaches for real-time anomaly detection in industrial control systems, in: Proceedings of the 6th ACM on Cyber-Physical System Security Workshop, CPSS 20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 230–29. doi:10.1145/3384941.3409588. URL https://doi.org/10.1145/3384941.3409588

[4] D. I. Urbina, J. A. Giraldo, A. A. Cardenas, N. O. Tippenhauer, J. Valente, M. Faisal, J. Ruths, R. Candell, H. Sandberg, Limiting the impact of stealthy attacks on industrial control systems, in: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, ACM, 2016, pp. 1092–1105.

[5] S. Athalye, C. M. Ahmed, J. Zhou, A tale of two testbeds: A comparative study of attack detection techniques in cps, in: A. Rashid, P. Popov (Eds.), Critical Information Infrastructures Security, Springer International Publishing, Cham, 2020, pp. 17–30.

[6] C. M. Ahmed, J. Zhou, Challenges and opportunities in cyberphysical systems security: A physics-based perspective, IEEE Security Privacy 18 (6) (2020) 14–22. doi:10.1109/MSEC.2020.3002851.

[7] F. Pasqualetti, F. Dorfler, F. Bullo, Attack detection and identification in cyber-physical systems, IEEE Transactions on Automatic Control 58 (2013) 2715–2729.

[8] C. M. Ahmed, A. Sridhar, M. Aditya, Limitations of state estimation based cyber attack detection schemes in industrial control systems, in: IEEE Smart City Security and Privacy Workshop, CPSWeek, 2016.

[9] C. Kwon, W. Liu, I. Hwang, Security analysis for cyber-physical systems against stealthy deception attacks, in: American Control Conference (ACC), 2013, pp. 3344–3349.

[10] Y. Mo, E. Garone, A. Casavola, B. Sinopoli, False data injection attacks against state estimation in wireless sensor networks, in: Decision and Control (CDC), 2010 49th IEEE Conference on, 2010, pp. 5967–5972.

[11] C. Murguia, J. Ruths, Characterization of a cusum model-based sensor attack detector, in: 55th IEEE Conference on Decision and Control Conference (CDC), 2016.

[12] V. R. Palleti, Y. C. Tan, L. Samavedham, A mechanistic fault detection and isolation approach using kalman filter to improve the security of cyber physical systems, Journal of Process Control 68 (2018) 160–170.

[13] S. Adepu, A. Mathur, Using process invariants to detect cyber attacks on a water treatment system, in: J.-H. Hoepman, S. Katzenbeisser (Eds.), ICT Systems Security and Privacy Protection, Springer International Publishing, Cham, 2016, pp. 91–104.

[14] H. Choi, W.-C. Lee, Y. Aafer, F. Fei, Z. Tu, X. Zhang, D. Xu, X. Deng, Detecting attacks against robotic vehicles: A control invariant approach, in: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 801–816. doi:10.1145/3243734.3243752.
URL https://doi.org/10.1145/3243734.3243752

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

[15] I. Y. Garrett, R. M. Gerdes, On the efficacy of model-based attack detectors for unmanned aerial systems, in: Proceedings of the Second ACM Workshop on Automotive and Aerial Vehicle Security, AutoSec 20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 11–14. doi:10.1145/3375706.3380555.
URL https://doi.org/10.1145/3375706.3380555

[16] C. M. Ahmed, C. Murguia, J. Ruths, Model-based attack detection scheme for smart water distribution networks, in: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, ASIA CCS 17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 101–113. doi:10.1145/3052973.3053011.
URL https://doi.org/10.1145/3052973.3053011

[17] P. Gunathilaka, D. Mashima, B. Chen, Softgrid: A software-based smart grid testbed for evaluating substation cybersecurity solutions, in: Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy, CPS-SPC 16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 113–124. doi:10.1145/2994487.2994494.
URL https://doi.org/10.1145/2994487.2994494

[18] B. Green, A. Lee, R. Antrobus, U. Roedig, D. Hutchison, A. Rashid, Pains, gains and plcs: Ten lessons from building an industrial control systems testbed for security research, in: 10th USENIX Workshop on Cyber Security Experimentation and Test (CSET 17), USENIX Association, Vancouver, BC, 2017.
URL https://www.usenix.org/conference/cset17/ //workshop-program/presentation/green

[19] P. P. Biswas, H. C. Tan, Q. Zhu, Y. Li, D. Mashima, B. Chen, A synthesized dataset for cybersecurity study of iec 61850 based substation, in: 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2019, pp. 1–7.

[20] Ã. L. Perales Gomez, L. Fernandez Maimo, A. Huertas Celdran, F. J. Garcia Clemente, C. Cadenas Sarmiento, C. J. Del Canto Masa, R. Mendez Nistal, On the generation of anomaly detection datasets in industrial control systems, IEEE Access 7 (2019) 177460–177473.

[21] iTrust, iTrust Datasets.
URL https://itrust.sutd.edu.sg/itrust-labs_datasets/

[22] V. R. Palleti, V. Kurian, S. Narasimhan, R. Rengaswamy, Actuator network design to mitigate contamination effects in water distribution networks, Computers & Chemical Engineering 108 (2018) 194–205.

[23] J. Slay, M. Miller, Lessons learned from the maroochy water breach, Springer 620 US, Boston, MA (2008) 73–82.

[24] I. C. 2014, Ics-mm201408: May-august 2014, Report no., U.S. Department of Homeland Security-Industrial Control Systems-Cyber Emergency Response Team, Washington, D.C. Available online at https://ics-cert.us-cert.gov.

[25] C. M. Ahmed, V. R. Palleti, A. P. Mathur, Wadi: A water distribution testbed for research in the design of secure cyber physical systems, in: Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, CySWATER 17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 25–28.

[26] R. Qadeer, C. Murguia, C. M. Ahmed, J. Ruths, Multistage downstream attack detection in a cyber physical system, in: Computer Security, Springer, 2017, pp. 177–185.

[27] C. Murguia, J. Ruths, Characterization of a cusum model-based sensor attack detector, in: 2016 IEEE 55th Conference on Decision and Control (CDC), 2016, pp. 1303–1309. doi:10.1109/CDC.2016.7798446.

[28] F. Pasqualetti, F. Dörfler, F. Bullo, Attack detection and identification in cyber-physical systems, IEEE transactions on automatic control 58 (11) (2013) 2715–2729.

[29] C. Murguia, J. Ruths, Cusum and chi-squared attack detection of compromised sensors, in: 2016 IEEE Conference on Control Applications (CCA), 2016, pp. 474–480. doi:10.1109/CCA.2016.7587875.

[30] J. Dressel, H. Farid, The accuracy, fairness, and limits of predicting recidivism, Science Advances 4 (1). arXiv:https://advances.sciencemag.org/content/4/1/eaao5580.full.pdf, doi:10.1126/sciadv.aao5580.
URL https://advances.sciencemag.org/content/4/1/eaao5580

[31] Y. Wang, Z. Xu, J. Zhang, L. Xu, H. Wang, G. Gu, Srid: State relation based intrusion detection for false data injection attacks in scada, in: M. Kutyłowski, J. Vaidya (Eds.), Computer Security - ESORICS 2014, Springer International Publishing, Cham, 2014, pp. 401–418.

[32] CNN, Florida city hacking incident into the water treatment system, https://edition.cnn.com/2021/02/08/us/oldsmar-florida-hack-water-poison/index.html, year = 2021.

APPENDIX

Shahid, M. O., Ahmed, C. M., Palleti, V. R., & Zhou, J. (2022).
Curse of system complexity and virtue of operational invariants: machine learning based system modeling and attack detection in CPS.
In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-8). IEEE.
https://doi.org/10.1109/dsc54232.2022.9888940

TABLE IV: Description of attacks launched on WADI testbed

Attack Identifier	Start time Start Index	End time End Index	Duration (minutes)	Attack description	
1	9/10/17 19:25:00 5101	9/10/17 19:50:16 6601	25.16	Motorized valve 1_MV_001 is maliciously turned on, this causes an overflow on primary tank	
2	10/10/17 10:24:10 59051	10/10/17 10:34:00 59641	9.50	Flow Indication Transmitter 1_FIT_001 is tuned off, a false reading is seen by PLC for 1_FIT_001. This will turn chemical dosing pump on while leaving the water level in primary tank constant. Consequently the attacker is increasing the level of chemicals inside water.	
3-4	10/10/17 10:55:00 60901	10/10/17 11:24:00 62641	29.0	Stealthy attack. Attacker aims to drain elevated reservoir 2_LT_002. This is done controlling manipulating tank level draining and filling speed. 1_AIT_001 Moreover the attacker changes the reading seen by water quality sensor, this causes the raw water tank drain.	
5	10/10/17 11:30:40 63041	10/10/17 11:44:50 63891	14.10	Turn off valves to consumers 2_MCV_101, 2_MCV_201, 2_MCV_301, 2_MCV_401, 2_MCV_501, 2_MCV_601 consumers will receive no more water.	
6	10/10/17 13:39:30 70771	10/10/17 13:50:40 71441	11.10	Turn on maliciously 2_MCV_101, 2_MCV_201.	
7	10/10/17 14:48:17 74898	10/10/17 14:59:55 75596	11.38	Supply contaminated water to the Elevated Reservoir tank by setting 1_AIT_002 to 6 to drain primary grid because of contamination. at the same time open 2_MV_003	
8	10/10/17 17:40:00 85201	10/10/17 17:49:40 85784	9.40	Maliciously open 2_MCV_007 in order to produce water leakage before water reaches consumers.	
9	11/10/17 10:55:00 147301	11/10/17 10:56:27 147388	1.27	Turn on 1_P_006 maliciously to cause pipe burst.	
10	11/10/17 11:17:54 148675	11/10/17 11:31:20 149481	14.26	Damage 1_MV_001 and raw water pump to drain Elevated Reservoir tank.	
11	11/10/17 11:36:31 149792	11/10/17 11:47:00 150421	11.29	Maliciously turn on 2_MCV_007 from 0% to 50% opening in order to produce leakage	
12	11/10/17 11:59:00 151141	11/10/17 12:05:00 151501	6.0	Maliciously turn on 2_MCV_007 from 0% to 100% opening with an increment of 10% opening at different times	
13	11/10/17 12:07:30 108451	11/10/17 12:10:52 108653	3.22	Reducing Booster set point pressure, this causes intermittent water supply to consumers	
14	11/10/17 12:16:00 152161	11/10/17 12:25:36 152737	9.36	stop chemical dosing to the raw water which is supplied to the primary grid tank tank	
15	11/10/17 15:26:30 163591	11/10/17 15:37:00 164221	11.30	Stealthy attack performed on 2_LT_002 to drain the elevated water tank	